Dynamic dimension reduction for financial applications

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)

eingereicht an der
Wirtschaftswissenschaftlichen Fakultät
Humboldt-Universität zu Berlin

von
MSc. Sergey Nasekin

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Wirtschaftswissenschaftlichen Fakultät:
Prof. Dr. Christian D. Schade

Gutachter:
1. Prof. Dr. Wolfgang Karl Härdle, Humboldt-Universität zu Berlin
2. Prof. Yingxing Li, Ph.D., Xiamen University

Tag des Kolloquiums: 20. Januar 2017

Abstract

Over the recent years, there have been a significant increase in financial data availability. On the other hand, financial markets have experienced sharp and often unforeseen changes in their dynamics. This tendency has caused the need for risk modeling approaches addressing both high dimensionality problem and accustoming for dynamic non-Gaussian structure.

The primary aim of this dissertation is to propose several risk modeling approaches which allow for simultaneous dimension reduction and dynamic structures in three setups: 1) asset allocation and hedging, 2) stochastic surface modeling and 3) systemic risk determination.

Proposed models demonstrate good performance when compared to existing approaches for risk modeling and introduce new flexible ways to detect extreme risks and anomalies on financial markets as well as methods for their modeling and management.

Key words: dimension reduction, trading strategies, dynamic factor models, systemic risk, factor copula, value-at-risk

Zusammenfassung

In den letzten Jahren gab es ein drastisches Wachstum in verfügbaren Finanzdaten. Finanzmärkte haben starke und oft nicht ganz vorhersagbare Änderungen ihrer Dynamik erlebt. Diese Tendenz hat dazu geführt, dass die Methoden der Risikomodellierung sowohl das Problem der hohen Dimensionalität als auch dynamische nicht-Gaußsche Strukturen behandeln müssen.

Das Ziel dieser Dissertation ist es, Methoden der Risikomodellierung vorzuschlagen, die gleichzeitig Reduzierung der Dimensionalität und dynamische Struktur in drei Anwendungen erlauben: 1) Asset Allocation und Hedging, 2) stochastische Modellierung von multivariaten Prozessen, 2) Messung der systemischen Risiken.

Die vorgeschlagenen Methoden demonstrieren gute Ergebnisse im Vergleich mit den existierenden Methoden der Risikomodellierung und führen neue Verfahren zur Erkennung der extremen Risiken und Anomalien auf Finanzmärkten sowie zur deren Management.

Schlagwörter: Reduzierung der Dimensionalität, Handelsstrategie, dynamische Faktormodelle, systemisches Risiko, Faktor-Copula, Value at Risk

Contents

List of Figures

List of Tables

Abbreviations

AIC	Akaike criterion
ARCH	Autoregressive conditional heteroscedasticity
ARMA	Autoregressive moving average
B&H	Buy-and-hold
BCBS	Basel Committee on Banking Supervision
CARA	Constant absolute risk aversion
CDO	Collateral debt obligation
CF	Cornish-Fisher
CVaR	Conditional Value-at-Risk
DSFM	Dynamic semiparametric factor model
GARCH	Generalized autoregressive conditional heteroscedasticity
GBM	Geometric Brownian motion
HQ	Hannan-Quinn criterion
ES	Expected shortfall
ETF	Exchange-traded fund
FSB	Financial Stability Board
ICA	Independent component analysis
IV	Implied volatility
LETF	Leveraged exchange-traded fund
LQR	Lasso quantile regression
MANIG	Multivariate affine normal inverse Gaussian distribution
MLE	Maximum likelihood estimation
NIG	Normal inverse Gaussian distribution
QR	Quantile regression
RMSPE	Root mean squared prediction error
SC	Schwarz criterion
SIFI	Systemically important financial institution
TEDAS	Tail-event driven asset allocation
VaR	Value-at-Risk
VIX	CBOE Volatility Index

Introduction

Over the recent decades have seen a sharp increase in the volume of financial data. Such dramatic increase poses opportunities as well as challenges for quantitative data analysis. As mentioned by Fan et al. (2013), new statistical thinking and computational methods are required to handle the challenges of big data. The challenges increase even more when a shift in data characteristics is considered. It is a well-known fact that financial time series exhibit such properties as non-stationarity, clustering of volatility and heavy-tailed distributions. The conditional evolution of variance has been documented in many research studies, most important being Engle (1982), Engle (2002). The assumption of dynamic univariate and multivariate conditional moments can be generalized to higher orders, as mentioned by Jondeau and Rockinger (2003).

Evidence for dynamic distributions should be accustomed for in various applications of financial modeling and analysis such as risk analysis, asset allocation, volatility modeling. In this work a dynamic distribution approach is introduced which accounts for time-varying higher moments in the asset portfolio distribution and allows to perform better assessment of portfolio risk. This technique is applied as the underlying model within a novel tail event asset allocation (TEDAS) approach applying a tail-event hedging strategy which improves asset allocation and gives superior performance with respect to market benchmarks.

Natural high dimensionality of financial data makes necessary dimension reduction which retains important information about the underlying processes and discards noise making computation and analysis feasible. As noticed in Fan et al. (2013), dimension reduction and variable selection play pivotal roles in analyzing high-dimensional data. Dynamic dimension reduction can be achieved through approximation of high-dimensional processes in lower-dimensional linear spaces. This approach is used in Chapter 2 of this study to construct dynamic implied volatility surfaces as part of the analysis of option market anomalies.

Factor representation can assist in modeling stochastic processes both in the one-dimensional case and in higher dimensions. In Chapter 3 of this study, a factor copula approach is applied to construct a systemic risk measure for a system of financial institutions. Systemic risk is an important aspect of economic risk which played a

significant role in the events of the financial crisis in 2007-2008. Its efficient definition quantification and definition is a topic of active research and attention of regulatory authorities, as mentioned by Basel Committee on Banking Supervision (2013). The factor copula approach allows to solve the high dimensionality problem introducing a factor dynamics for variables and reducing computation effort, as mentioned by Krupskii and Joe (2013).

Variable selection or regularization assists in selecting relevant parameters in sparse settings. A wide range of regularization tools have been proposed recently, e.g., in Tibshirani (1996), Fan and Li (2001) and Zou and Hastie (2005). In the TEDAS strategy mentioned above, a Lasso quantile regression with non-positive constraints is used to select assets relevant for hedging purposes out of a large universe of securities available. Such an approach simultaneously allows to mitigate risk and increase financial returns of portfolio strategies.

A dynamic semiparametric approach is used to implement lower-dimensional space factor representations of implied volatility surfaces. Implied volatility analysis is applied to options on leveraged exchange-traded funds (LETFs) which exhibit an anomalous condition of implied volatility discrepancy. The dynamic modeling study confirms the findings of uniform confidence bands' analysis which studies statistical significance of a moneyness scaling adjustment over time. This adjustment uses specific assumptions on the underlying stochastic model of the asset return process. While the standard Black-Scholes framework is assumed for the purposes of the statistical analysis, a semi-closed form expression for the expected integrated variance in the case of stochastic volatility is also derived. Furthermore, a Monte-Carlo technique to estimate this value demonstrates that the expectation of the integrated stochastic variance assumes an U-shape resembling the implied volatility smile.

This study is structured as follows. In Chapter 1, the novel TEDAS asset allocation strategy is described. This strategy simultaneously allows to hedge against extreme negative events in the market and model asset portfolios in a dynamic non-Gaussian setting. The technical details of the TEDAS strategy are outlined. The Lasso shrinkage method and quantile regression as well as their joint implementation - the Lasso quantile regression estimator are presented. This technique is used to perform asset selection in the first step of the TEDAS strategy. Portfolio allocation methods based on multi-moment utility and risk measures are introduced. These measures are made dynamic via the concept of time-varying conditional distribution. A simulation analysis of the non-positive Lasso quantile regression estimator is provided and the TEDAS strategy is empirically tested. It is shown to demonstrate better out-of-sample performance relative to other common asset allocation techniques such as Markowitz risk-return optimization and a naive allocation approach.

In Chapter 2, the statistical properties of the moneyness scaling transformation are studied. This transformation adjusts the moneyness coordinate of the implied

volatility smile in an attempt to remove the discrepancy between the IV smiles for levered and unlevered ETF options. It is reasonable to assume that the option contracts on leveraged and unleveraged ETFs should have similar implied volatility structures having the same source of randomness. However, these structures generally do not correspond according to empirical observation. Assuming the Black-Scholes framework, Leung and Sircar (2015) propose the method of moneyness scaling which apparently removes the discrepancy. However, using bootstrap uniform confidence bands for a robust non-parametric estimator of implied volatility, it is demonstrated that there remains a possibility that the implied volatility smiles are not the same, even after moneyness scaling has been performed.

This presents possible arbitrage opportunities on the (L)ETF market which can be exploited by traders. An empirical data application shows that there are indeed such opportunities in the market which result in risk-free gains for the investor. A dynamic "trade-with-the-smile" strategy based on a dynamic semiparametric factor model is presented. This strategy utilizes the dynamic structure of implied volatility surface allowing out-of-sample forecasting and information on unleveraged ETF options to construct theoretical one-step-ahead implied volatility surfaces. Additionally, a semi-analytic and a simulation-based estimation approach are proposed for incorporating stochastic volatility into the moneyness scaling method. This approach allows to infer the "expected integrated variance smile" from the data.

In Chapter 3, a network-based factor copula approach is proposed to study systemic risk in a network of systemically important financial institutions (SIFIs). A network-based approach is also employed to identify the most "connected" institutions and the results are compared to the factor copula model outcome. The motivation for this research is determined by the growing importance of accurate measurement of global systemic risk. A market-based approach is used and it is found that the resulting classification of SIFIs differs from the official "bucket" system by the Basel Committee on Banking Supervision. The findings of the copula method are largely confirmed by the network-based technique. The proposed approach aims to identify extreme tail dependence between the institutions and therefore accurately captures the level of risk during financial crises. At the same time, dimension reduction effect is achieved because the dynamics of portfolio is modeled via a common market factor and an idiosyncratic factor.

Chapter 1

TEDAS: Tail Event Driven ASset Allocation

Introduction

Portfolio selection and risk management are important concepts in quantitative finance and applied statistics. Their applications often deal with portfolio assets' correlation structure estimation. The correlation structure, or, more generally, the dependence across assets is a main component of the portfolio allocation problem since it determines the level of risk in the investment position. However, the correlation is not informative on the distributional details of the portfolio. It does not specify the dependence between assets at different quantiles, but refers to relations with respect to their mean values, which may be weak, while relations or even dependence in tails or, more broadly, quantiles or expectiles, may be significant. Assets, which have negative correlation when the markets are stable, may exhibit positive correlation during volatile periods. Modelling tail dependence is a more informative and flexible approach to hedging and portfolio allocation. Bassett et al. (2004) propose the so-called α-risk measure minimization approach which is compatible with pessimistic Choquet preferences and is formulated as a quantile regression problem. Other authors directly consider more traditional downside risk measures for portfolio optimization: Alexander and Baptista (2002) study a mean-VaR model for portfolio selection, Gaivoronski and Pflug (2004) compare Value-at-Risk, mean-variance and conditional VaR portfolio optimization approaches.

Statistical estimation in regression problems and multivariate distribution moment computation used for asset allocation is often inaccurate or even infeasible if the number of covariates p is very large, possibly larger than the number of observations n. To address this issue, penalization techniques have been proposed: the L_1 penalized least squares method defined as LASSO (Least Absolute Shrinkage

and Selection Operator) was introduced by Tibshirani (1996). Another common penalty function, SCAD, was proposed by Fan and Li (2001); the so-called "elastic net" which is a linear combination of L_1 and L_2 penalties, was introduced in Zou and Hastie (2005). Chang and Tsay (2010) apply the L_1-penalized normal log-likelihood method to estimate a high-dimensional covariance matrix. An application to asset allocation is discussed by Brodie et al. (2009) and Fan et al. (2009) who impose the Lasso constraint on portfolio weights. Wu et al. (2014) study the constrained index tracking problem in stock market without short sales using the non-negative Lasso approach which imposes an additional non-negativity constraint on parameters.

In this study, an alternative approach is proposed, which is based on a two-step procedure which uses the adaptive Lasso-penalized quantile technique to identify tail risk and multivariate dynamic higher-moment portfolio optimization to obtain portfolio weights. This new approach is defined as a "Tail Event Driven ASset allocation (TEDAS) strategy" which is based on a given benchmark asset and "hedging assets" and includes estimation of quantile dependence between assets' returns in the case of high dimensionality with $p > n$. It accounts for possible dynamic, time-varying distributional characteristics of the portfolio.

The issue of asset choice, which is crucial in asset allocation, is discussed, and the strategy is implemented empirically using hedge funds as "hedging assets" in portfolio optimization. In earlier, such as Lintner (1983) and more recent literature, as Cvitanić et al. (2003), Favre and Galeano (2002), Giamouridis and Vrontos (2007), Lhabitant and Learned (2002), McFall Lamm (1999), McFall Lamm (2003), the use of hedge funds as portfolio assets along with conventional securities such as stocks or bonds has been advocated because they provide superior risk-adjusted returns as well as diversification benefits due to their dynamic nature, non-equity related strategies and other features. It has become an established fact that most hedge fund strategies exhibit asymmetric return patterns characterized by negative skew and excess kurtosis due to using leverage and financial derivatives. Therefore a successful portfolio allocation and risk measurement procedure should be able to match higher moments of the portfolio distribution such as skewness and kurtosis.

This chapter is structured as follows. The next section outlines the technical details of the TEDAS strategy. The Lasso shrinkage method and quantile regression as well as their joint implementation - the Lasso quantile regression estimator are presented. This technique is used to perform asset selection in the first step of the TEDAS strategy. Portfolio allocation methods based on multi-moment utility and risk measures are introduced. These measures are made dynamic via the concept of time-varying conditional distribution. Section 2 provides a simulation analysis of the non-positive Lasso quantile regression estimator as well as empirical testing of the TEDAS strategy, which is shown to demonstrate better out-of-sample performance relative to the market performance and naive equal-weight allocation.

1.1 TEDAS - Tail Event Driven ASset Allocation Strategy

1.1.1 Description

The TEDAS strategy selects portfolio assets oppositely related to a benchmark ("core") asset/index in the lower and positively in the upper tail of its conditional distribution. "Tail events" are defined as the co-movements between random financial returns at different quantiles of their distributions. Dependence of variables based on conditional quantiles is measured via the quantile regression approach. Variable selection via Lasso penalization allows to reduce the complexity of the dependence structure and make it computationally feasible in the next step: portfolio selection.

Given that several "hedging" assets have been chosen, the problem of optimal portfolio composition arises. The assets' returns may exhibit asymmetric return patterns characterized by negative skew and excess kurtosis. The traditional risk-return analysis does not address these facts and the idea of higher-moment optimization is a natural alternative here. Therefore one has to modify the objective risk measure which would make the allocation more tractable in the case of non-normality.

In general, the two-step TEDAS procedure can be outlined as follows:

1. select the relevant hedging assets using the tail event approach;
2. perform portfolio selection using an asset allocation criterion with appropriate distributional characteristics.

1.1.2 First Step: Asset Selection

Tail event detection is done via the high-dimensional Lasso-penalized quantile regression approach. This technique allows to exclude irrelevant covariates, making the model parsimonious and reducing its prediction error. The Lasso estimator, see Tibshirani (1996), was first proposed for a linear model in the least-squares framework. It is used to avoid model overfitting by imposing the L_1-penalty on the coefficients and shrinking them to zero. Therefore it becomes possible to estimate coefficients of a high-dimensional design matrix, where the number of covariates p may be much larger than the number of observations n.

Quantile regression estimation provides conditional quantile functions which describe the relation between response and regressors for some quantile level $\tau \in (0,1)$: consider a random sample from some distribution $\{(X_i, Y_i); i = 1, ..., n\}$,

$X_i \in \mathbb{R}^p$, $Y_i \in \mathbb{R}$. Given the piecewise linear loss function $\rho_\tau(u) = |u|\{\tau - \mathbf{I}(u < 0)\}$, the quantile regression estimator is the solution to the convex optimization problem:

$$\hat{\beta}_\tau = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i^\top \beta); \tag{1.1}$$

the conditional quantile function is given by $q_\tau(x) \stackrel{\text{def}}{=} F^{-1}_{y|x}(\tau|x) = x^\top \beta(\tau)$. The L_1-penalized quantile regression (LQR) estimator is then constructed as follows:

$$\begin{aligned} \hat{\beta}_{\tau,\lambda} = \quad & \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i^\top \beta) \\ & \text{subject to:} \quad t - \|\beta\|_1 \geq 0. \end{aligned} \tag{1.2}$$

where t is the size constraint on $\|\beta\|_1$. In unrestricted form, (1.2) is equivalent to:

$$\hat{\beta}_{\tau,\lambda} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i^\top \beta) + \lambda \|\beta\|_1. \tag{1.3}$$

As first noted by Barrodale and Roberts (1974) and later by Koenker and Bassett (1978), (1.1) is equivalent to a linear program which is also the case for the LQR. There is a correspondence between λ and t which depends on the data X, Y and can be illustrated by the duality of (1.2), see Osborne et al. (2000). The choice of the regularization parameter λ is crucial for the Lasso estimator. It controls the level of penalization and the resulting shrinkage. The methods to select λ optimally, such as cross-validation, generalized cross-validation or information criteria are discussed, for instance, in Tibshirani (1996).

In the context of the LQR, it is especially relevant to consider the case of high-dimensional sparse models where the overall number of regressors p is very large, possibly much larger than the sample size n, but the number of significant regressors for each conditional quantile of interest is at most q, which is smaller than the sample size, that is, $q = o(n)$. A number of general regularity conditions needed for the derivation of the Lasso-penalized quantile regression estimator are usually introduced, as in Belloni and Chernozhukov (2011).

Existing critiques of the Lasso model point out that the resulting estimates are generally asymptotically biased and are not consistent, see, e. g., Hastie et al. (2011), Bühlmann and van de Geer (2011). Several approaches have been proposed to address this problem which modify the penalty and appropriately select the penalization parameter λ_n to achieve good statistical properties. For example, the adaptive Lasso approach introduced by Zou (2006) uses a re-weighted L_1 penalty, where the weights can be obtained from any root-n-consistent LQR estimator; the "clipped" SCAD penalty introduced by Fan and Li (2001), does not excessively penalize large values of β.

It is essential to determine the regularization parameter λ so that the resulting estimator retains "good" properties such as asymptotic normality or variable selection consistency. Traditional procedures to determine λ, such as cross-validation and information criteria, have several drawbacks. As p increases with the growth of the sample size, the number of potential models goes to infinity very quickly and there is no guarantee that, for instance, K-fold cross-validation will provide a choice of λ with a proper rate. Belloni and Chernozhukov (2011) suggest using a data-driven procedure to select the penalty level. They also derive a LQR estimator which is consistent at a rate satisfied by the choice of λ. The estimator is formulated as follows:

$$\hat{\beta}_{\tau,\lambda} = \arg\min_{\beta\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(Y_i - X_i^\top \beta) + \frac{\lambda\sqrt{\tau(1-\tau)}}{n}\sum_{i=j}^{p} \widehat{w}_j|\beta_j|, \tag{1.4}$$

where $\widehat{w}_j^2 \stackrel{\text{def}}{=} n^{-1}\sum_{i=1}^n X_{ij}^2$; λ is chosen via a simulation scheme, see Belloni and Chernozhukov (2011). Such a choice of λ leads to optimal rates of convergence of the estimator (1.4) and yields good statistical properties.

The TEDAS tail events' approach requires to estimate only non-positive LQR coefficients, which correspond to the hedging assets X negatively related to the benchmark Y in the lower tail of its conditional distribution. Such assets tend to have positive returns when the benchmark asset return is negative at specified quantile levels τ and vice versa. This allows to perform asset allocation more precisely and hedge benchmark asset tail events, when the downside risk is especially high. So the final LQR estimator used for the purpose of the TEDAS strategy is as follows:

$$\begin{aligned}\hat{\beta}_{\tau,\lambda} = \quad & \arg\min_{\beta\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(Y_i - X_i^\top \beta) + \frac{\lambda\sqrt{\tau(1-\tau)}}{n}\sum_{i=j}^{p} \widehat{w}_j|\beta_j| \\ & \text{subject to:} \quad \beta_j < 0.\end{aligned} \tag{1.5}$$

The non-positivity requirement amounts to adding one more constraint in the linear program formulation of the LQR problem, see 4.1.

1.1.3 Second Step: Portfolio Selection

Asset Allocation Methods

Many portfolio managers rely on the Markowitz (mean-variance or risk-return) rule which combines assets into an "efficient" portfolio offering risk-adjusted target returns. Assuming a dynamic distribution approach, consider a random return process $\{R_t\}_{t\in\mathscr{T}}$ on the filtered probability space $(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}_t\}_{t\in\mathscr{T}})$, $\mathscr{T}$ is some

ordered index set, with realizations $r_t \overset{\text{def}}{=} R_t(\omega)$. Also assume that $r_t|\mathscr{F}_{t-1} \sim F_t$ where F_t denotes the conditional cumulative distribution function of r_t given the information up to the moment $t-1$, $\mathscr{F}_{t-1}$. The conditional expectation $\mathsf{E}(\cdot|\mathscr{F}_{t-1})$ is denoted as $\mathsf{E}_{t-1}(\cdot)$.

Risk-return optimization is based on four inputs: the weights of total funds invested in each security w_{it}, $i = 1,\ldots,d$, the expected returns μ, volatilities (standard deviations) σ_{it} associated with each security and covariances σ_{ijt}, $j = 1,\ldots,d; i \neq j$ between returns. Portfolio weights w_{it} are obtained from the quadratic optimization problem for the portfolio variance $\sigma^2_{P,t}(w_t)$:

$$\begin{aligned} \underset{w_t \in \mathbb{R}^d}{\text{minimize}} \quad & \sigma^2_{P,t}(w_t) \overset{\text{def}}{=} w_t^\top \Sigma_t w_t \\ \text{subject to} \quad & \mu_{P,t}(w_t) = r_T, \\ & w_t^\top 1_d = 1, \\ & w_{i,t} \geq 0 \end{aligned} \tag{1.6}$$

where $\Sigma_t \overset{\text{def}}{=} \mathsf{E}_{t-1}\{(r_t - \mu)(r_t - \mu)^\top\}$ is the covariance matrix for r_{it}, $\mu_{P,t}(w_t) \overset{\text{def}}{=} w_t^\top \mu$, $\mu \overset{\text{def}}{=} \mathsf{E}_{t-1}(r_t)$ is the portfolio mean, r_T is the "target" return for the portfolio assigned by the investor. The Markowitz rule simultaneously solves two problems: diversification and asset allocation. Diversification reduces specific risk; asset allocation allows to combine assets so that the portfolio risk can be lowered while the expected returns are not necessarily reduced. The exact shape of the curve of possible allocations depends on correlations between assets: the smaller the correlation, the smaller the risk of the portfolio. Therefore one prefers to find assets that offer an acceptable return while being less than perfectly correlated or even negatively correlated.

Although normality of returns is not a requirement for using the mean-variance optimization strategy, only the normal distribution is uniquely defined by the first two moments. A natural further generalization has been to extend this assumption to the higher moments of the portfolio, see Jurczenko and Maillet (2006). The so-called mean-variance-skewness-kurtosis optimization criterion assumes that the asset return distributions belong to a four-parameter family of probability distributions which allow for finite mean, variance, skewness and kurtosis. This choice reflects the fact of distributional skewness and fat-tailedness for financial time series. In the context of portfolio optimization, skew-Student distributions were discussed by Jondeau and Rockinger (2009), the generalized hyperbolic distribution was applied by Ghalanos et al. (2015).

The mean-variance-skewness-kurtosis allocation approach can be formulated as follows: consider an investor who allocates the portfolio to maximize the expected utility $\mathsf{E}_{t-1}\{U(W_t)\}$ over the end-of-period wealth W_t. Given certain assumptions about the "well-behavedness" of a von Neumann-Morgenstern utility function $U(\cdot)$,

see Jurczenko and Maillet (2006), one can write the investor's problem as:

$$\max_{w_t \in \mathbb{R}^d} \mathsf{E}_{t-1}\{U(W_t)\}, \text{ s.t. } \mu_{P,t}(w_t) = r_T, w_t^\top 1_d = 1, w_{i,t} \geq 0, \tag{1.7}$$

where, after ignoring higher-than-fourth moment terms, it is readily shown that, in the case of a risk-averse investor with an exponential (CARA, constant absolute risk aversion) utility $U(W_t) = -\exp(-\eta W_t)$, where η is the coefficient of risk aversion,

$$\mathsf{E}_{t-1}\{U(W_t)\} \approx -\exp(-\eta \overline{W}_t)\left(1 + \frac{\eta^2}{2}\sigma_{W_t}^2 - \frac{\eta^3}{3!}S_{W_t} + \frac{\eta^4}{4!}K_{W_t}\right). \tag{1.8}$$

Asset allocation based on tail risk minimization takes an extreme downside risk measure as the optimization criterion. The Value-at-Risk (VaR) measures the maximum portfolio loss given confidence level α: it is an α-quantile of the probability distribution of r with a cumulative distribution function F:

$$q_\alpha \stackrel{\text{def}}{=} \text{VaR}_\alpha \stackrel{\text{def}}{=} \inf\{r : F(r) \geq \alpha\}. \tag{1.9}$$

A modification of VaR via the Cornish-Fisher (CF) expansion improves its precision adjusting estimated quantiles for non-normality. As discussed by Favre and Galeano (2002), VaR based only on volatility underestimates portfolio risk. The CF expansion, see Abramowitz and Stegun (1965), approximates the quantile of an arbitrary random variable Y with mean μ, variance σ^2 and the cumulative distribution function F_Y via a standard normal variate $z_\alpha \stackrel{\text{def}}{=} \Phi^{-1}(\alpha)$ and higher moments. Let y_α be α-quantile, $F_Y(y_\alpha) = \alpha$, then the CF approximation yields: $y_\alpha \simeq \mu - \sigma q_\alpha$, where the fourth-order CF approximation provides the following expression for q_α:

$$q_\alpha = z_\alpha + (z_\alpha^2 - 1)\frac{S}{6} + (z_\alpha^3 - 3z_\alpha)\frac{K}{24} - (2z_\alpha^3 - 5z_\alpha)\frac{S^2}{36}, \tag{1.10}$$

where S and K are skewness and excess kurtosis, respectively.

To incorporate asymmetry explicitly into the allocation procedure, one needs to calculate the *portfolio* skewness and kurtosis making optimization over higher moments possible and thereby refining risk assessment. The time-varying portfolio skewness $S_{P,t}$ and kurtosis $K_{P,t}$ are defined, respectively, for the $(d \times d^2)$ co-skewness and $(d \times d^3)$ co-kurtosis matrices M_t^3 and M_t^4:

$$M_t^3 \stackrel{\text{def}}{=} \mathsf{E}_{t-1}\{(r_t - \mu)(r_t - \mu)^\top \otimes (r_t - \mu)^\top\} \tag{1.11}$$

$$M_t^4 \stackrel{\text{def}}{=} \mathsf{E}_{t-1}\{(r_t - \mu)(r_t - \mu)^\top \otimes (r_t - \mu)^\top \otimes (r_t - \mu)^\top\}, \tag{1.12}$$

where each element of M_t^3, M_t^4 is given as

$$M_{ijk,t}^3 = \mathsf{E}_{t-1}\{(r_{i,t} - \mu_i)(r_{j,t} - \mu_j)(r_{k,t} - \mu_k)\}$$
$$M_{ijkl,t}^4 = \mathsf{E}_{t-1}\{(r_{i,t} - \mu_i)(r_{j,t} - \mu_j)(r_{k,t} - \mu_k)(r_{l,t} - \mu_l)\}\}.$$

The moments of portfolio returns can be then computed in a tractable way for a given vector of portfolio asset weights w_t:

$$S_{P,t}(w_t) = \frac{w_t^\top M_t^3 (w_t \otimes w_t)}{(w_t^\top \Sigma_t w_t)^{3/2}}$$
$$K_{P,t}(w_t) = \frac{w_t^\top M_t^4 (w_t \otimes w_t \otimes w_t)}{(w_t^\top \Sigma_t w_t)^2}.$$

The CF-VaR expansion (1.10) in the dynamic multivariate case takes the form (with $w = w_t$):

$$q_\alpha(w_t) = z_\alpha + (z_\alpha^2 - 1)\frac{S_{P,t}(w_t)}{6} + (z_\alpha^3 - 3z_\alpha)\frac{K_{P,t}(w_t)}{24} - (2z_\alpha^3 - 5z_\alpha)\frac{S_{P,t}(w_t)^2}{36},$$

which can be used in the new VaR optimization procedure to obtain optimal weight allocation for portfolio assets w_t. This allocation approach seeks to minimize the tail risk of the portfolio, taking into account extreme events. Normally this optimization problem is not convex, compared to the Markowitz case and typically exhibits multiple local minima. Moreover, its has exponential computational complexity in the dimension of w_t. Therefore, one needs a feasible dimension reduction scheme to make the computations tractable.

Being a popular measure of risk, VaR suffers from several drawbacks. It is often unstable and difficult to work with numerically when losses are not normally distributed. As noted by Artzner et al. (1999), the VaR measure is not *coherent*: it fails to satisfy the so-called *sub-additivity property* and violates the principle that "a merger does not create extra risk". Another shortcoming of VaR is that it provides no information on the extent of the losses that might be suffered beyond the threshold amount indicated by this measure. The alternative measure which is simultaneously coherent and robust is the *conditional Value-at-Risk* or CVaR, first proposed by Rockafellar and Uryasev (2000). The issue of portfolio moments' incorporation into the CVaR risk measure for the portfolio is solved by approximating the CVaR via the Cornish-Fisher expansion.

Proposition 1.1.3.1. Given the confidence level $\alpha > 0.5$ and the investment horizon T, the $\text{CVaR}_{1-\alpha}$ can be approximated with the Cornish-Fisher expansion as follows:

$$\text{CVaR}_\alpha = -\frac{1}{1-\alpha} q_\alpha^*(w_t)\sigma_{P,t}(w_t)\sqrt{T},$$

where

$$q_\alpha^*(w_t) = \left\{1 + \frac{S_{P,t}(w_t)}{6} z_\alpha + \frac{K_{P,t}(w_t)}{24}(z_\alpha^2 - 1) - \frac{S_{P,t}^2(w_t)}{36}(2z_\alpha^2 - 1)\right\}\varphi(z_\alpha). \tag{1.13}$$

Proof. See Appendix 4.2. □

It is straightforward to formulate the asset allocation problem as minimization of the Cornish-Fisher VaR/CVaR tail risk measure. Here the simple VaR measure is used in the empirical study for backtesting purposes. The optimal portfolios w_t are calculated from:

$$\min_{w_t \in \mathbb{R}^d} \text{CVaR}_\alpha(w_t), \text{ s.t. } \mu_{P,t}(w_t) = r_T, w_t^\top 1_d = 1, w_{i,t} \geq 0, \tag{1.14}$$

where

$$\text{CVaR}_\alpha(w_t) \stackrel{\text{def}}{=} -\frac{1}{1-\alpha} q_\alpha^*(w_t)\sigma_{P,t}(w_t), \tag{1.15}$$

with $q_\alpha^*(w_t)$ as in (1.13), $S_P = S_{P,t}(w_t)$, $K_P = K_{P,t}(w_t)$, $T = 1$.

Time-Varying Conditional Distribution

The assumption of univariate as well as multivariate normality of financial returns has been challenged in numerous research on the topic. Financial returns are often known to be skewed and leptokurtic. Furthermore, financial time series often exhibit such phenomena as volatility clustering or conditional heteroscedasticity. Sample autocorrelation functions for the squared returns of selected hedge funds' indices have been estimated and it was found that they imply persistence in the variance of the returns' series. It is logical to assume that conditional moments higher than variance are also not constant in time. The phenomenon of conditional heteroscedasticity may be extended to the conditional skewness and kurtosis leading to conditional *heterocliticity* and *heterokurticity*. In Table 1.1 descriptive statistics, results of statistical tests for univariate as well as multivariate normality and tests for heteroscedasticity (ARCH test), heterocliticity (Bera-Lee) and heterokurtosis (Bera-Zuo) are given for hedge funds' indices focused on 3 different regions: Japan, North America and Europe. Univariate normality assumption is rejected in all three cases; multivariate normality also does not hold. Moreover the three tests for variation in second, third and fourth conditional moments in all but one case result in rejecting the corresponding null hypothesis. The Bera-Zuo test for heterokurtosis does not reject constant conditional kurtosis for the North America Fixed Income Fund of Funds index.

Therefore an appropriate model for portfolio returns has to account for univariate and multivariate dynamic asymmetry as well as leptokurticity of variables. A number of approaches addressing the problem of multivariate dynamic volatility modelling, among the most well-known are GARCH-type BEKK, see Engle and Kroner (1995), dynamic conditional correlation (DCC), see Engle (2002), orthogonal GARCH, see Alexander (2001), have been restricted to explaining the dynamics of first and second multivariate conditional moments assuming multivariate normality. The issue of both dynamic returns' moments and asymmetry as well as leptokurticity has been discussed in non-elliptical conditional distributions modeling framework. The class of generalized hyperbolic (GH) distributions allows to

Table 1.1: Statistics on monthly returns of Eurekahedge hedge funds' indices

	Japan Multi-Strategy		North America Fixed Income		Europe Arbitrage	
Univariate statistics						
Moments						
Mean	0.008	(0.002)	0.006	(0.001)	0.003	(0.001)
Std	0.021	(0.001)	0.017	(0.001)	0.018	(0.001)
Skew	2.067	(0.191)	−1.595	(0.191)	0.553	(0.191)
Kurt	7.782	(0.381)	5.710	(0.381)	9.358	(0.381)
Normality tests						
JB	533.775	(0.000)	294.089	(0.000)	610.407	(0.000)
KS	0.503	(0.000)	0.473	(0.000)	0.485	(0.000)
Omnibus	82.773	(0.000)	43.761	(0.000)	171.079	(0.000)
Dynamic conditional moments' tests						
ARCH	11.227	(0.000)	34.966	(0.000)	26.592	(0.000)
Bera-Lee	48.469	(0.000)	36.475	(0.000)	40.783	(0.000)
Bera-Zuo	203.723	(0.000)	20.149	(0.166)	421.847	(0.000)
Multivariate statistics						
Test						
Omnibus	326.226	(0.000)				
Mardia	301.199	(0.000)				
Henze-Zirkler	9.862	(0.000)				

* Calculations made for monthly data; standard errors and p-values are given in parentheses
** Univariate summary statistics Mean, Std, Skew and Kurt denote the mean, the standard deviation, the skewness and the excess kurtosis of returns, respectively; JB, KS and Omnibus stand for the Jarque-Bera (Jarque and Bera, 1980), Kolmogorov-Smirnov and Omnibus (see below) statistics for the test of the null hypothesis of a normal distribution. ARCH, Bera-Lee and Bera-Zuo stand for the test statistics of the ARCH test by information matrix tests for testing variation in second, third and fourth conditional moments by Bera and Lee (1993) and Bera and Zuo (1996), respectively. The null hypothesis is absence of conditional heteroscedasticity, heterocliticity and heterokurticity, respectively.
*** Omnibus, Mardia and Henze-Zirkler stand for the statistics proposed by Doornik and Hansen (2008), Mardia (1970) and Henze and Zirkler (1990), respectively, testing the null of multivariate normality

model heavier tails than the Gaussian law as well as provides for asymmetric shape fitting. The properties of the GH distributions are discussed in Barndorff-Nielsen and Blaesild (1981). Jondeau and Rockinger (2009) propose using the multivariate time-varying conditional skew-T distribution to model portfolio higher moments. In Ghalanos et al. (2015) the so-called autoregressive conditional GH distribution is fitted to financial returns' data allowing to explicitly calculate multivariate conditional portfolio moments.

A sub-class of the generalized hyperbolic law, the normal-inverse Gaussian (NIG) distribution, was introduced in mathematical finance research by Barndorff-Nielsen (1997). The density of the NIG$(\mu, \delta, \alpha, \beta)$ with shape parameters α and β con-

trolling skewness and kurtosis as well as location and scale parameters μ and δ, respectively, is written as follows:

$$f_{NIG}(x) = \frac{\alpha\delta}{\pi} \exp\left\{\delta\sqrt{\alpha^2-\beta^2} + \beta(x-\mu)\right\} \frac{K_1\left\{\alpha\sqrt{\delta^2+(x-\mu)^2}\right\}}{\sqrt{\delta^2+(x-\mu)^2}}, \quad (1.16)$$

where $0 \leq |\beta| \leq \alpha$, $\delta > 0$, K_1 is the modified Bessel function of the third kind and order 1. The NIG distribution has the "semi-heavy" tails property, which ensures that the NIG distribution is appropriate for fitting financial data which do not exhibit extreme leptokurticity but are still heavier-tailed than the Gaussian law.

An proper distributional model for portfolio returns would address the non-normality assumption and produce the dynamic moments in (1.11), (1.12) as well as possess computational feasibility. The class of generalized hyperbolic (GH) distributions without further assumptions has the drawback that the margins of the multivariate GH (MGH) distribution are not mutually independent for some choice of the dependence structure $\Sigma = AA^\top$. The multivariate *affine* GH (MAGH) distribution, discussed by Schmidt et al. (2006), models margins and dependency separately. It holds that $Y \sim \text{MAGH}(\lambda, \alpha, \beta, \mu, \Sigma)$ if

(i) $X = (X_1, \ldots, X_d)^\top$, $X_i \sim \text{GH}(\lambda_i, \alpha_i, \beta_i, 0, 1)$, $i = 1, \ldots, d$, X_i independent

(ii) $Y = AX + \mu$, the matrix $AA^\top$ is positive definite.

An appropriate choice for the matrix A is made via the so-called *independent component analysis (ICA)*. The ICA technique assumes a random signal X which is generated by another random vector of independent components (ICs) $s = (s_1, \ldots, s_d)$, s_i statistically independent, $i = 1, \ldots, d$ and a quadratic mixing matrix A, both assumed to be unknown. The ICA approach separates source signals s from a set of mixed signals X without or with very little aid of information about s or the mixing process A. Both A and s are estimated by maximizing the non-Gaussianity of linear combinations of X; the fixed-point FastICA algorithm with cubic convergence for this problem is described in Hyvärinen et al. (2001).

The model for portfolio returns is given by a dynamic multivariate affine normal-inverse Gaussian (MANIG) distribution as follows:

$$r_t = m_t + \varepsilon_t, \text{ and } r_t|\mathscr{F}_{t-1} \sim \text{MANIG}(m_t, \Sigma_t, \omega_t), \quad (1.17)$$

where $\omega_t = (\omega_{1t}, \ldots, \omega_{dt})^\top$ and $\omega_{it} = (\alpha^*_{it}, \beta^*_{it})^\top$, $i = 1, \ldots, d$, conditional mean m_t and conditional covariance matrix Σ_t.

Under the assumptions of the model, the underlying process has the specification $z_t \stackrel{\text{def}}{=} D_t^{-1/2} s_t$, where $z_t = (z_{1t}, \ldots, z_{dt})^\top$, $z_{it} \sim NIG(\alpha_{it}, \beta_{it}, 0, 1/\sqrt{d_{it}})$, z_{it}, z_{jt-s} independent, $\forall j \neq i$, $\forall s$; the conditional variance matrix $D_t \stackrel{\text{def}}{=} \mathsf{E}(s_t s_t^\top|\mathscr{F}_{t-1}) =$

$\text{diag}(d_{1t},\ldots,d_{dt})$ so that each of the independent components $s_{it} \sim NIG(\alpha_{it}/\sqrt{d_{it}}, \beta_{it}/\sqrt{d_{it}}, 0, 1)$ which satisfies the independence assumption (i). It is also assumed that $\varepsilon_t = As_t$, A being the mixing matrix, $\mathsf{E}(z_t|\mathscr{F}_{t-1}) = 0$, then it follows that $\mathsf{E}(s_t|\mathscr{F}_{t-1}) = 0$ and $\Sigma_t = \mathsf{E}(r_t r_t^\top) = AD_tA^\top$.

The properties of the multivariate NIG distribution allow us to determine the higher-moment dynamic matrices M_t^3 and M_t^4 defined in (1.11) and (1.12). The skew and shape parameters β_{it} and α_{it} together determine the higher moments of the NIG distribution such as skewness and kurtosis. It is convenient to reparametrize the model for z_{it} to two parameters in control of skew and shape, respectively: $\xi_{it} = \beta_{it}/\alpha_{it}$, $\nu_{it} = \sqrt{d_{it}\left(\alpha_{it}^2 - \beta_{it}^2\right)}$. Furthermore, following Jondeau and Rockinger (2009), the skew and shape parameters are assumed to follow asymmetric GARCH-like dynamics:

$$\nu_{i,t} = a_{i,0} + a_{i,1}^{-}|s_{i,t-1}|N_{i,t-1} + a_{i,1}^{+}|s_{i,t-1}|P_{i,t-1} + a_{i,2}\nu_{i,t-1} \tag{1.18}$$

$$\xi_{i,t} = b_{i,0} + b_{i,1}^{-}s_{i,t-1}N_{i,t-1} + b_{i,1}^{+}s_{i,t-1}P_{i,t-1} + b_{i,2}\xi_{i,t-1}, \tag{1.19}$$

where $N_{i,t} = \mathbf{I}(s_{i,t} \leq 0)$, $P_{i,t} = 1 - N_{i,t}$. The shape parameter $\nu_{i,t}$ is related to the absolute value of lagged signals: a large shock $s_{i,t-1}$ is expected to impact the fat-tailedness of the distribution irrespective of its sign. The dynamics of the skew parameter $\xi_{i,t}$ is related to signed shocks because the asymmetry is affected by both the sign and the size of $s_{i,t-1}$. Finally, the dynamics of $\nu_{i,t}$ and $\xi_{i,t}$ are assumed to be autoregressive to capture possible persistence of conditional skewness and kurtosis.

Estimation is done via the maximum likelihood method on an equation-by-equation basis due to marginal independence. A logistic map technique is employed to constrain the parameters in the autoregressive equations (1.18) and (1.19). In Table 1.3, the maximum-likelihood estimates for the portfolio of 10 hedge funds most frequently chosen by the TEDAS strategy over different benchmark indices are shown. The hedge funds used for estimation are shown in Table 1.2. It can be seen that, for instance, volatility persistence is very weak for most of the independent components. Moreover, most of the parameters in the conditional skewness dynamics are significant over the majority of the components, which confirms the existence of portfolio distributional asymmetry persistence in time.

From the estimation results in Table 1.3 it can be seen that almost all significant parameters a_1^- and a_1^+ are negative. This suggests that, irrespective of the sign, a large shock produces a subsequent distribution with fatter tails. It is also observed that $|a_1^-| < |a_1^+|$, implying that a positive shock increases the distribution's tails more so than a negative shock does. As regards the skew parameter $\xi_{i,t}$, b_1^- have positive signs and so negative shocks tend to decrease the subsequent distributional skewness and, therefore, increase the probability of a negative shock in the following period. A large positive shock, however, can both decrease and increase the subsequent skewness (b_1^+ positive or negative) and, hence, increase or decrease the probability of losses in the next period.

Fund name	Domicile	Strategy
All benchmarks		
R&C Hedge FIM	Brazil	Mixed Asset
Pure Heart Value Investment Fund	Cayman Islands	Long/Short Equity
Epic Wisdom Advanced	Cayman Islands	Equity Market Neutral
FIM LP Reims	Brazil	Mixed Asset
SZITIC Star Rock I	China	Long Bias
Wada Capital Japan Trust	Cayman Islands	Equity Market Neutral
BNY Mellon Enhanced Coefficient	Ireland	Global Macro
Cambrian Asia Ltd.	Cayman Islands	Developed Asia Equity
Everyoung Opportunities	Cayman Islands	Emerging Markets
Fraternity Hedge	Cayman Islands	Multi Strategy

Table 1.2: Top 10 overall selected hedge funds for all benchmark indices

Once the dynamics of $v_{i,t}$ and $\xi_{i,t}$ have been obtained, one readily calculates the conditional dynamic autoregressive skewness and excess kurtosis sk^s_{it}, $kurt^s_{it}$ for each $s_{i,t}$:

$$sk^s_{it} = 3\frac{\xi_{i,t}}{\sqrt[4]{d_{it}(1-\xi^2_{i,t})}\sqrt{\alpha_{it}}}$$
$$kurt^s_{it} = \frac{12\xi^2_{i,t}+3}{\alpha_{it}\sqrt{d_{it}(1-\xi^2_{i,t})}}.$$

Then the expressions for M^3_t, M^4_t in (1.11) and (1.12) are given by

$$M^3_t = AM^3_{s_t}(A\otimes A)^\top$$
$$M^4_t = AM^4_{s_t}(A\otimes A\otimes A)^\top,$$

where

$$M^3_{s_t} = \mathsf{E}_{t-1}(s_{i,t}s_{j,t}s_{k,t}) = \sum_{r=1}^{p} d_{ir,t}d_{jr,t}d_{kr,t}sk^s_{rt}$$
$$M^4_{s_t} = \mathsf{E}_{t-1}(s_{i,t}s_{j,t}s_{k,t}s_{l,t})$$
$$= \sum_{r=1}^{p} d_{ir,t}d_{jr,t}d_{kr,t}d_{lr,t}kurt^s_{rt} + \sum_{r=1}^{p}\sum_{s\neq r}\psi_{rs,t},$$

where $\psi_{rs,t} = d_{ir,t}d_{jr,t}d_{ks,t}d_{ls,t} + d_{ir,t}d_{js,t}d_{kr,t}d_{ls,t} + d_{is,t}d_{jr,t}d_{kr,t}d_{ls,t}$, $D^{1/2}_t = (d_{ij,t})_{i,j=1,\ldots,p}$.

Table 1.3: Skew and shape dynamics' parameter estimates

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
ω	0.042^{***}	0.873^{***}	0.482^{***}	0.893^{***}	0.812^{***}	0.456^{***}	0.218^{***}	0.000	0.000	0.000
α_1	0.000^{***}	0.127^{***}	0.009^{***}	0.002^{***}	0.010^{***}	0.001^{***}	0.002^{***}	0.593^{***}	0.000^{***}	0.932^{***}
β_1	0.730^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.000	0.000^{***}	0.407^{***}	0.355^{***}	0.015^{***}
a_0	-0.060^{***}	0.035	-0.033^{***}	-0.019	-0.005^{***}	-0.020	0.025^{***}	-0.014^{***}	0.113^{***}	-0.005
a_1^-	0.016	0.005^{***}	-0.033^{***}	0.011	-0.006^{***}	0.039^{**}	0.099^{***}	-0.056^{***}	0.229^{***}	-0.048^{*}
a_1^+	0.497	-0.071	-0.095^{***}	-0.011	0.137^{***}	-0.034	-0.073^{***}	0.031	-0.125^{***}	-0.083^{*}
a_2	-0.001	0.972	0.990^{***}	0.716	0.731^{***}	0.896	1.797^{***}	0.891^{***}	0.803^{***}	0.413
b_0	-2.836^{***}	-2.905^{***}	-2.220^{***}	-2.568^{***}	-3.655^{***}	-4.081^{***}	-3.326^{***}	-5.389^{***}	-2.725^{***}	-4.665^{***}
b_1^-	0.295	0.496^{***}	0.499^{***}	-0.640^{***}	-1.525^{***}	0.346^{***}	-0.038^{***}	-0.088^{***}	-0.342^{***}	-0.154^{***}
b_1^+	-3.067	-0.917^{***}	-0.913^{***}	-0.218^{***}	-1.867^{***}	-0.679^{***}	0.160^{***}	-0.382^{***}	-0.333^{***}	-0.484
b_2	-0.260^{***}	-0.997^{***}	-1.004^{***}	-0.387^{***}	0.379^{***}	0.076^{***}	-0.846	0.746^{***}	-0.392^{***}	-1.301
$Factor_{LL}$	519.03	2226.98	476.37	1268.70	864.16	1181.75	1245.75	1546.32	3156.41	1986.12
$Model_{LL}$	66643									

[*] Estimates of dynamic skew and shape equations' parameters
[**] Standard errors obtained using a numeric Hessian procedure
[***] The $*$, $**$ and $***$ denote significance at the 10%, 5% and 1% levels, respectively

1.2 Simulation Study and Data Analysis

1.2.1 Simulation study

In the following simulation study, constrained (non-positive and non-negative) and unconstrained LQR estimates are compared according to several accuracy measures. Two linear model designs are considered: the first with $\beta^{(1)} = (-5,-5,-5,-3,-1,-0.5,0,...,0)$ and the second with $\beta^{(2)} = (5,5,5,3,1,0.5,0,...,0)$. The following setup holds for both models: $X_i \sim \mathrm{N}(0,\Sigma)$, $i = 1,\ldots,n$; the regressors are correlated with $\Sigma_{jk} = 0.5^{|j-k|}$, $j,k = 1,\ldots,p$, $n = 100$, $p = 500$; ε_i are independent, $\varepsilon_i \sim \mathrm{N}(0,\sigma^2)$. The true dimension of the model is thus $q = 6$ (the number of negative-signed variables). Three levels of noise are considered: $\sigma = 0.1, 0.5, 1$. Model selection is performed via a simulation procedure, as outlined in Belloni and Chernozhukov (2011). The accuracy of the model selection is assessed with 4 accuracy criteria: estimate of the true model dimension (Est), standardized L_2-norm (Dev), least angle (Angle) and empirical risk (Risk), see 4.3 for details.

The results of the simulation analysis under three levels of noise and three quantile indices: $\tau = 0.1$, $\tau = 0.5$ and $\tau = 0.9$ are shown in Tables 1.4 and 1.5. The constrained LQR method demonstrates better accuracy than the unconstrained estimator. Other accuracy criteria which measure estimates' precision such as Dev and Angle are either similar or better on average for the constrained model. The Risk measure capturing the empirical risk is sometimes slightly better for the unconstrained model. However, for a TEDAS strategy, accuracy in terms of the correct estimated dimension is more relevant as an indicator of model adequacy, because the dimension and the sign matter more than the estimates' magnitude.

Table 1.4: Constrained non-positive and unconstrained LQR: accuracy criteria results

Accuracy Crit. and Model		Noise Levels and Quantile Indices								
		σ=0.1			σ=0.5			σ=1		
		$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
Est	**Constrained**	6.09	5.98	6.04	6.25	5.90	6.18	6.28	5.55	6.24
		(0.93)	(0.23)	(0.92)	(0.99)	(0.39)	(0.99)	(1.20)	(0.56)	(1.24)
	Unconstrained	6.69	6.50	6.85	6.94	6.50	6.98	6.91	6.26	7.14
		(1.61)	(1.14)	(1.58)	(1.71)	(1.30)	(1.66)	(1.81)	(1.43)	(1.77)
Dev	**Constrained**	0.05	0.02	0.06	0.09	0.07	0.09	0.13	0.11	0.13
		(0.10)	(0.02)	(0.10)	(0.06)	(0.02)	(0.06)	(0.05)	(0.03)	(0.06)
	Unconstrained	0.08	0.02	0.06	0.12	0.08	0.10	0.17	0.13	0.16
		(0.14)	(0.02)	(0.10)	(0.09)	(0.03)	(0.07)	(0.09)	(0.04)	(0.08)
Angle	**Constrained**	0.99	1.00	0.99	1.00	1.00	1.00	0.99	0.99	0.99
		(0.04)	(0.00)	(0.04)	(0.02)	(0.00)	(0.02)	(0.02)	(0.00)	(0.02)
	Unconstrained	0.99	1.00	0.99	0.99	1.00	0.99	0.98	0.99	0.99
		(0.05)	(0.01)	(0.04)	(0.03)	(0.01)	(0.03)	(0.03)	(0.01)	(0.03)
Risk	**Constrained**	0.17	0.06	0.18	0.30	0.23	0.30	0.41	0.35	0.41
		(0.32)	(0.06)	(0.33)	(0.21)	(0.07)	(0.22)	(0.17)	(0.08)	(0.18)
	Unconstrained	0.20	0.06	0.14	0.31	0.23	0.28	0.44	0.35	0.40
		(0.35)	(0.06)	(0.26)	(0.22)	(0.08)	(0.19)	(0.22)	(0.10)	(0.18)

Unconstrained: unconstrained Lasso quantile regression; Constrained: Lasso quantile regression with non-positive constraints

Standard deviations are given in brackets

Number of Monte-Carlo replications is 1000

Table 1.5: Constrained non-negative and unconstrained LQR: accuracy criteria results

Accuracy Crit. and Model		Noise Levels and Quantile Indices								
		σ=0.1			σ=0.5			σ=1		
		$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
Est	**Constrained**	6.05	6.00	6.07	6.16	5.90	6.19	6.21	5.58	6.32
		(0.92)	(0.21)	(0.90)	(0.98)	(0.39)	(1.03)	(1.18)	(0.55)	(1.20)
	Unconstrained	6.69	6.47	6.84	6.90	6.47	7.04	7.01	6.22	7.05
		(1.61)	(1.11)	(1.48)	(1.74)	(1.31)	(1.70)	(1.83)	(1.43)	(1.79)
Dev	**Constrained**	0.06	0.02	0.05	0.09	0.07	0.09	0.13	0.11	0.13
		(0.10)	(0.02)	(0.10)	(0.07)	(0.02)	(0.06)	(0.06)	(0.03)	(0.05)
	Unconstrained	0.07	0.02	0.05	0.11	0.08	0.10	0.16	0.13	0.15
		(0.13)	(0.02)	(0.09)	(0.08)	(0.03)	(0.07)	(0.08)	(0.04)	(0.07)
Angle	**Constrained**	0.99	1.00	0.99	0.99	1.00	0.99	0.99	1.00	0.99
		(0.04)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.00)	(0.02)
	Unconstrained	0.99	1.00	0.99	0.99	1.00	0.99	0.98	0.99	0.99
		(0.05)	(0.01)	(0.03)	(0.03)	(0.01)	(0.02)	(0.03)	(0.01)	(0.02)
Risk	**Constrained**	0.19	0.06	0.18	0.30	0.22	0.29	0.43	0.37	0.43
		(0.35)	(0.05)	(0.34)	(0.23)	(0.06)	(0.20)	(0.19)	(0.09)	(0.18)
	Unconstrained	0.18	0.06	0.14	0.30	0.22	0.28	0.43	0.36	0.41
		(0.33)	(0.05)	(0.24)	(0.22)	(0.07)	(0.17)	(0.20)	(0.09)	(0.17)

Unconstrained: unconstrained Lasso quantile regression; Constrained: Lasso quantile regression with non-negative constraints

Standard deviations are given in brackets

Number of Monte-Carlo replications is 1000

1.2.2 Empirical Testing of the TEDAS Strategy

Rolling window estimation

The TEDAS strategy is tested for four largest international stock markets: American S&P 500, Japanese Nikkei 225, British FTSE 100 and German DAX 30. Data on daily log-returns of 888 hedge funds in the period 20100205-20160203 are used. The data are obtained from the Thomson Reuters Lipper Investment Management Database. Additionally to the primary indices above, two additional benchmark indices are analyzed: the leveraged exchange-traded funds ProShares UltraPro S&P500 which tracks the S&P 500 index with the leverage coefficient +3 and ProShares UltraPro QQQ tracking NASDAQ 100 index with the same leverage coefficient. The LETF indices are high-performing compared to the other benchmarks and normally offer higher performance but also a larger amount of risk and increased possibility of extreme events. This would potentially make a tail-event hedging strategy such as TEDAS relevant in this case.

A rolling-window dynamic estimation of the TEDAS strategy is performed in order to conduct backtesting analysis of the dynamic Value-at-Risk measure afterwards as well as select the relevant hedging assets chosen with the highest frequency in rolling-window estimations. The analysis results are further used for static portfolio performance analysis.

The distribution model for r_t is assumed to be the MANIG model in (1.17). It is assumed that m_t follows AR(1) specification, conditional variances d_{it} are modeled according to the GARCH(1,1) approach. The TEDAS strategy can be specified as follows: let the initial wealth $W_0 = \$1$, at $t = l, \ldots, n$, where l is the length of the moving window, denote the the log-returns' empirical cumulative distribution function as $\widehat{F}_n(r_t)$ then at each time step t:

1. determine the benchmark return r_t, set $\tau_t = \widehat{F}_n(r_t)$,
2. solve (1.5) for $\hat{\beta}_{\tau_t,\lambda_n}$ using the observations $X \in \mathbb{R}^{t-l+1,\ldots,t\times p}$, $Y \in \mathbb{R}^{t-l+1,\ldots,t\times 1}$,
3. if $r_t < 0$, choose $X^{\mathscr{J}}_{t+1}$ for investing, where $\mathscr{J}$ is the index set corresponding to non-zero coefficients $\hat{\beta}_{\tau_t,\lambda_n}$ from Step 2; if $r_t > 0$, invest into the index r_{t+1},
4. if $X^{\mathscr{J}}_{t+1}$ are chosen, apply an asset allocation procedure to $X^{\mathscr{J}}$ to determine $\hat{w}_t$,
5. determine the realized portfolio wealth for $t+1$ depending on the outcome in Step 3: $W_{t+1} = W_t(1 + \hat{w}_t^\top X^{\mathscr{J}}_{t+1})$ if $X^{\mathscr{J}}_{t+1}$ are chosen or $W_{t+1} = W_t(1 + r_{t+1})$ if r_t is chosen.

The log-returns' empirical cumulative distribution function $\widehat{F}_n(r_t)$ is a continuous approximation of the empirical distribution function of the benchmark index log-returns. Here a kernel density estimator is used: $\widehat{F}_n(r) \stackrel{\text{def}}{=} \int_{-\infty}^{r} \hat{f}_n(u)\,du =$
$= \frac{1}{n}\sum_{i=1}^{n} H\{(r-r_i)/h\}$, $\hat{f}_n(x) \stackrel{\text{def}}{=} (1/nh)\sum_{i=1}^{n} K\{(r-r_i)/h\}$, $H(x) = \int_{-\infty}^{x} K(u)du$, $K(\cdot)$ is a kernel function, h is bandwidth. For the purpose of this study, the Gaussian kernel is used and the bandwidth is chosen according to the Silverman's rule of thumb, see Silverman (1998).

3 TEDAS investing strategies are constructed which are differentiated by the allocation procedure in Step 4:

1. TEDAS 1: Cornish-Fisher VaR optimization,
2. TEDAS 2: expected CARA utility optimization (assume $\eta = 10$),
3. TEDAS 3: Markowitz risk-return optimization.

2 benchmark strategies are additionally used to compare against TEDAS in terms of performance:

1. $1/p$: naive equal weights allocation for the whole universe of assets,
2. B&H: index buy-and-hold strategy.

We do not consider here conventional asset allocation approaches such as mean-variance or Choquet portfolio selection proposed by Bassett et al. (2004) which are not robust to the problem of high dimensionality in the case when $p > n$.

In Figures 1.1-1.2 four approaches to portfolio allocation are compared in terms of cumulative portfolio performance. TEDAS strategies mostly deliver better performance compared to the benchmark naive $1/p$ and B&H strategies assuming perfect financial markets (no transaction costs). Comparison of strategies in terms of risk and return via Sharpe ratios is given in Table 1.6. However, the assumption of zero transaction costs is often unrealistic and constant portfolio rebalancing will make TEDAS strategies perform worse than buy-and-hold rules.

Therefore, dynamic assessment of the TEDAS strategies can be regarded as a preliminary step for constructing static portfolios which can be compared to the market with zero transaction costs as a realistic assumption in the situation when no rebalancing takes place. Furthermore, dynamic estimation allows to assess the quality of Cornish-Fisher VaR model in terms of its relevance for the data. Backtesting results are portrayed in Figures 1.3, 1.4. Backtesting via the Christoffersen likelihood ratio test is performed to evaluate adequacy of the VaR model for particular benchmark

	T1	T2	T3	$1/p$	Buy&Hold
S&P500	0.20	0.34	0.43	−0.65	0.42
FTSE100	0.57	−0.03	0.51	−0.65	−0.09
DAX30	0.31	0.44	0.34	−0.65	0.16
NIKKEI225	0.00	−0.34	−0.20	−0.65	0.34
UPRO	0.13	0.33	0.25	−0.65	0.19
TQQQ	0.56	0.37	0.51	−0.65	0.29

Table 1.6: Sharpe ratios for different strategies and benchmark indices; dynamic estimation, zero transaction costs assumed

indices. The model can be accepted as correct on 1%-level in 4 out of 6 cases (for S&P500, DAX30, NIKKEI225 and TQQQ as benchmark indices).

Frequency analysis is performed to determine which hedge funds are most universal for protection against tail events. The histograms of the estimated dimension $\hat{q}$ for daily data are shown in Figure 1.5. The median estimated dimension is equal to 11 for most cases with the exception of FTSE 100 and DAX 30 benchmarks where it is equal to 8. In Figure 1.6 the frequency of the selected hedge funds is demonstrated. At the first glance, one of the funds is consistently chosen for all benchmark indices and another one is chosen in three cases as one of the two most frequently chosen. These insights are confirmed by the analysis in Table 1.7. The hedge fund R&C Hedge FIM with the domicile in Brazil is most frequently selected for 5 out of 6 index benchmarks and as fourth most frequent in the case of the NIKKEI 225 benchmark. Another Brazilian hedge fund, FIM LP Reims, is chosen as second most frequent for the American market indices (S&P 500, UPRO and TQQQ). Both of these hedge funds pursue mixed-asset investment strategies which presumably capture most portfolio diversification gains.

Static portfolio performance

Dynamic analysis outlined in the previous section does not include transaction costs which invalidate the performance results for strategies with frequent portfolio rebalancing. However this type of analysis may be regarded as backtesting analysis, which would yield candidates for static portfolio construction. First analysis similar to that for the results in Table 1.7 is performed. First 500 rolling-window estimations are used to obtain 5 most frequent hedge funds for each benchmark index. Then static portfolios are constructed from each set of 5 hedge funds and the corresponding benchmark index, respectively (6-dimensional portfolios). These (optimized) portfolios are held for the rest of data sample period. So out-of-sample performance of static portfolios can be compared to market index performance. As the portfolios are now static, transaction costs are not an issue.

Standard risk-return allocation is used to construct 2 types of optimized portfolios: high target return (higher risk) and maximum Sharpe ratio (tangency, lower risk) portfolios. Additionally, naive equal-weight portfolios are constructed and compared to the performance of the market (benchmark) index in terms of standard performance metrics: Sharpe ratios, Modigliani risk-adjusted performance ("M-squared") and Jensen's alpha.

The sample is divided into the training sample and test sample periods 20100205-20121220 and 20121221-20160203, respectively; 500 rolling-window LQR $\widehat{q}$ estimates are used to determine the hedge funds most frequently chosen as benchmark index hedges. Then performance of static portfolios described above is compared to market performance. This static portfolio composition method corresponds to the so-called core-satellite approach where passively and actively managed securities are combined together for possible performance gains. The results of this analysis are summarized in Table 1.8. The resulting portfolio dynamics is shown in Figures 1.7, 1.8.

It can be seen that the constructed portfolios outperform the market in all the cases except for one case of naive allocation with NIKKEI 225 as the benchmark index. The flexibility of the TEDAS approach allows to improve performance by choosing a high-performance index as the benchmark, such as leveraged ETFs ProShares UltraPro S&P 500 and QQQ.

In general, the empirical analysis demonstrates that the TEDAS two-step approach based on tail events identification and multi-moment dynamic portfolio selection shows superior performance compared to other asset allocation strategies proposed in the literature such as the traditional Markowitz optimization and pessimistic portfolio allocation by Bassett et al. (2004). It also allows to better reveal portfolio diversification opportunities on the global markets due to its robustness to the problem of high dimensionality.

1.3 Conclusion

The main methodological contribution of this study is to develop a two-step asset allocation strategy which identifies the tail risk of a benchmark asset and uses multi-moment dynamic portfolio selection technique to account for possible conditional non-normality of portfolio returns.

The TEDAS - Tail Event ASset Allocation strategy is based on the non-negative/non-positive Lasso adaptive quantile regression method which penalizes model complexity and shrinks a high-dimensional asset universe so that it is feasible for an appropriate asset allocation procedure. In the second step, the dynamic investor risk/utility measures such as the Cornish-Fisher conditional Value-at-Risk and four-

moment expected utility are developed to perform multi-moment portfolio selection. This procedure assumes neither joint nor marginal normality of assets' returns and allows to incorporate multivariate portfolio skewness and kurtosis statistics into the optimization process.

The TEDAS strategy captures both negative left tail events for the selected benchmark assets. It is tested for four major international markets and is shown to demonstrate superior out-of-sample performance compared to the market performance as well as naive equal-weight allocation.

Fund name	Frequency	Domicile	Strategy
S&P 500 benchmark			
R&C Hedge FIM	460	Brazil	Mixed Asset
FIM LP Reims	452	Brazil	Mixed Asset
Pure Heart Value Investment Fund	273	Cayman Islands	Long/Short Equity
SZITIC Star Rock I	224	China	Long Bias
Epic Wisdom Advanced	219	Cayman Islands	Equity Market Neutral
NIKKEI 225 benchmark			
Pure Heart Value Investment Fund	278	Cayman Islands	Long/Short Equity
Transtrend Dvsfd Trend Program	248	Netherlands	Managed Futures/CTAs
INFINITY Platinum FIM	240	Brazil	Mixed Asset
R&C Hedge FIM	229	Brazil	Mixed Asset
Wada Capital Japan Trust	209	Cayman Islands	Equity Market Neutral
FTSE 100 benchmark			
R&C Hedge FIM	341	Brazil	Mixed Asset
Fraternity Hedge	286	Cayman Islands	Multi Strategy
Wada Capital Japan Trust	250	Cayman Islands	Equity Market Neutral
Hayate Japan Equity Long-Short	234	Cayman Islands	Long/Short Equity
BNY Mellon Enhanced Coefficient	225	Ireland	Global Macro
DAX 30 benchmark			
R&C Hedge FIM	433	Brazil	Mixed Asset
Cambrian Asia Ltd.	259	Cayman Islands	Developed Asia Equity
Epic Wisdom Advanced	205	Cayman Islands	Equity Market Neutral
Everyoung Opportunities	205	Cayman Islands	Emerging Markets
Wada Capital Japan Trust	191	Cayman Islands	Equity Market Neutral
UPRO benchmark			
R&C Hedge FIM	469	Brazil	Mixed Asset
FIM LP Reims	442	Brazil	Mixed Asset
Pure Heart Value Investment Fund	278	Cayman Islands	Long/Short Equity
Epic Wisdom Advanced	226	Cayman Islands	Equity Market Neutral
SZITIC Star Rock I	224	China	Long Bias
TQQQ benchmark			
R&C Hedge FIM	489	Brazil	Mixed Asset
FIM LP Reims	346	Brazil	Mixed Asset
Pure Heart Value Investment Fund	279	Cayman Islands	Long/Short Equity
SZITIC Star Rock I	238	China	Long Bias
Perimeter Private 3 FIM	234	Brazil	Mixed Asset

Table 1.7: Top 5 selected hedge funds for each of 6 benchmarks

Table 1.8: Comparison of static portfolio performance with core-satellite approach

Benchmark index	Portfolios and performance measures											
	Max. return			Max. Sharpe ratio			Equal weights			Market		
	Sharpe	M^2	α	Sharpe	M^2	α	Sharpe	M^2	α	Sharpe	M^2	α
S&P 500	0.073	0.0003	0.0004	0.082	0.0003	0.0003	0.060	0.0001	0.0003	0.042	NA	NA
FTSE 100	0.057	0.0005	0.0003	0.059	0.0005	0.0004	0.033	0.0003	0.0001	−0.003	NA	NA
DAX 30	0.051	0.0004	0.0005	0.065	0.0005	0.0003	0.051	0.0004	0.0003	0.022	NA	NA
NIKKEI 225	0.051	0.0007	0.0001	0.053	0.0001	0.0002	−0.048	−0.0014	−0.0013	0.046	NA	NA
UPRO	0.075	0.0009	0.0005	0.080	0.0010	0.0004	0.070	0.0007	0.0003	0.040	NA	NA
TQQQ	0.078	0.0008	0.0004	0.086	0.0010	0.0005	0.067	0.0004	0.0003	0.052	NA	NA

* Portfolios are constructed as follows: (1) Max. return: a portfolio on a high-risk end of the Markowitz efficient frontier; (2) Max. Sharpe ratio: the tangent Markowitz portfolio maximizing the Sharpe ratio; (3) Equal weights: all portfolio components are taken with equal weights; (4) Market: the benchmark index

** Performance measures: (1) Sharpe: Sharpe ratio, defined as $\text{Sharpe} \stackrel{\text{def}}{=} (r_p - r_f)/\sigma_p$, where r_p is the portfolio return, r_f is a risk-free rate, σ_p is the portfolio standard deviation; (2) $M^2 \stackrel{\text{def}}{=} \sigma_m(r_p - r_f)/\sigma_p - (r_m - r_f)$, where r_m is the market return, σ_m is the market standard deviation; (3) $\alpha \stackrel{\text{def}}{=} r_p - \{r_f + \beta_p(r_m - r_f)\}$ is Jensen's alpha; here β_p is the portfolio beta as defined in the CAPM model; $\beta_p \stackrel{\text{def}}{=} \mathsf{Cov}(r_p, r_m)/\mathsf{Var}(r_m)$

*** All performance measures are averages over the test sample period 20121221-20160203; the risk-free rate is assumed to be zero; markets are benchmark indices for each case, respectively

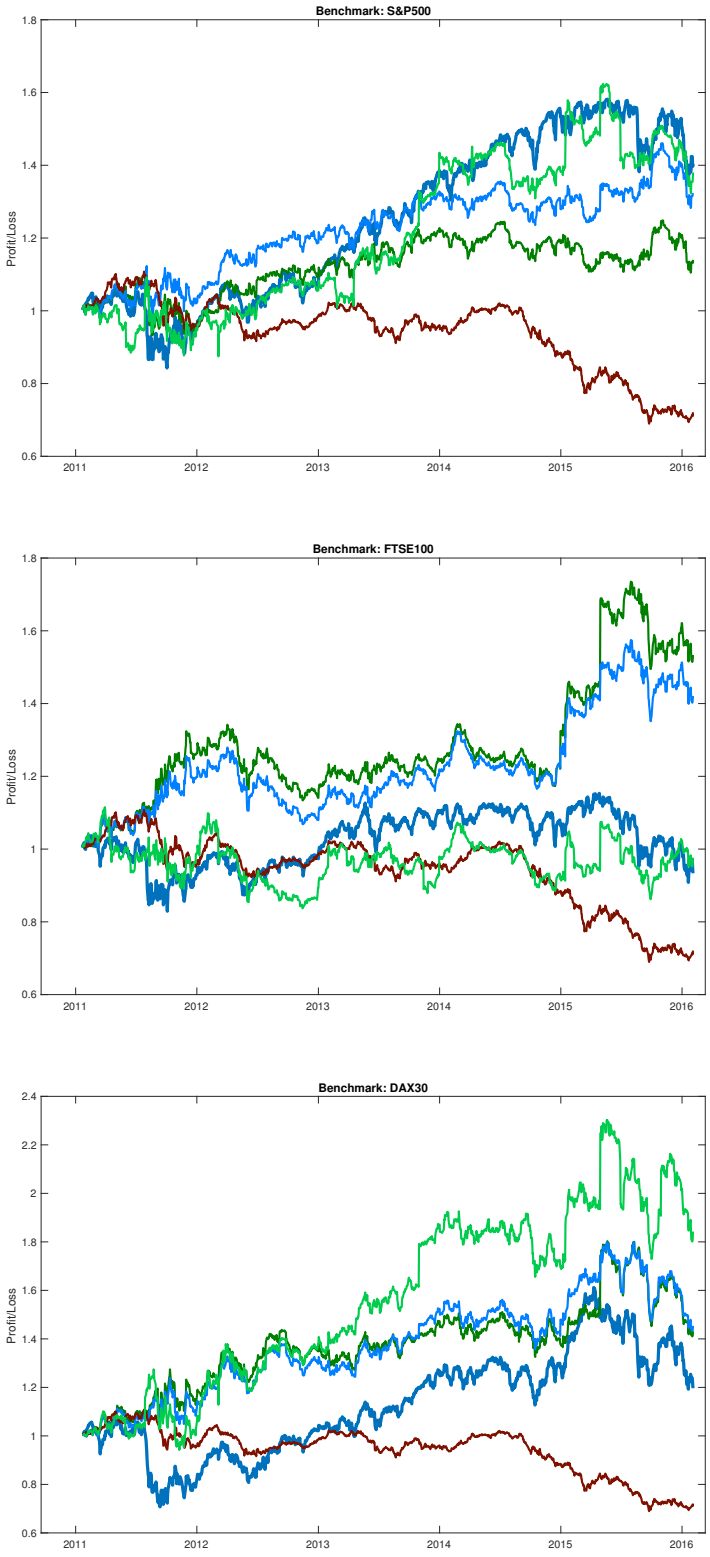

Figure 1.1: Cumulative portfolio wealth comparison: TEDAS 2, TEDAS 3, TEDAS 4, $1/p$; upper panel: S&P500 B&H, middle panel: FTSE100 B&H, lower panel: DAX30 B&H

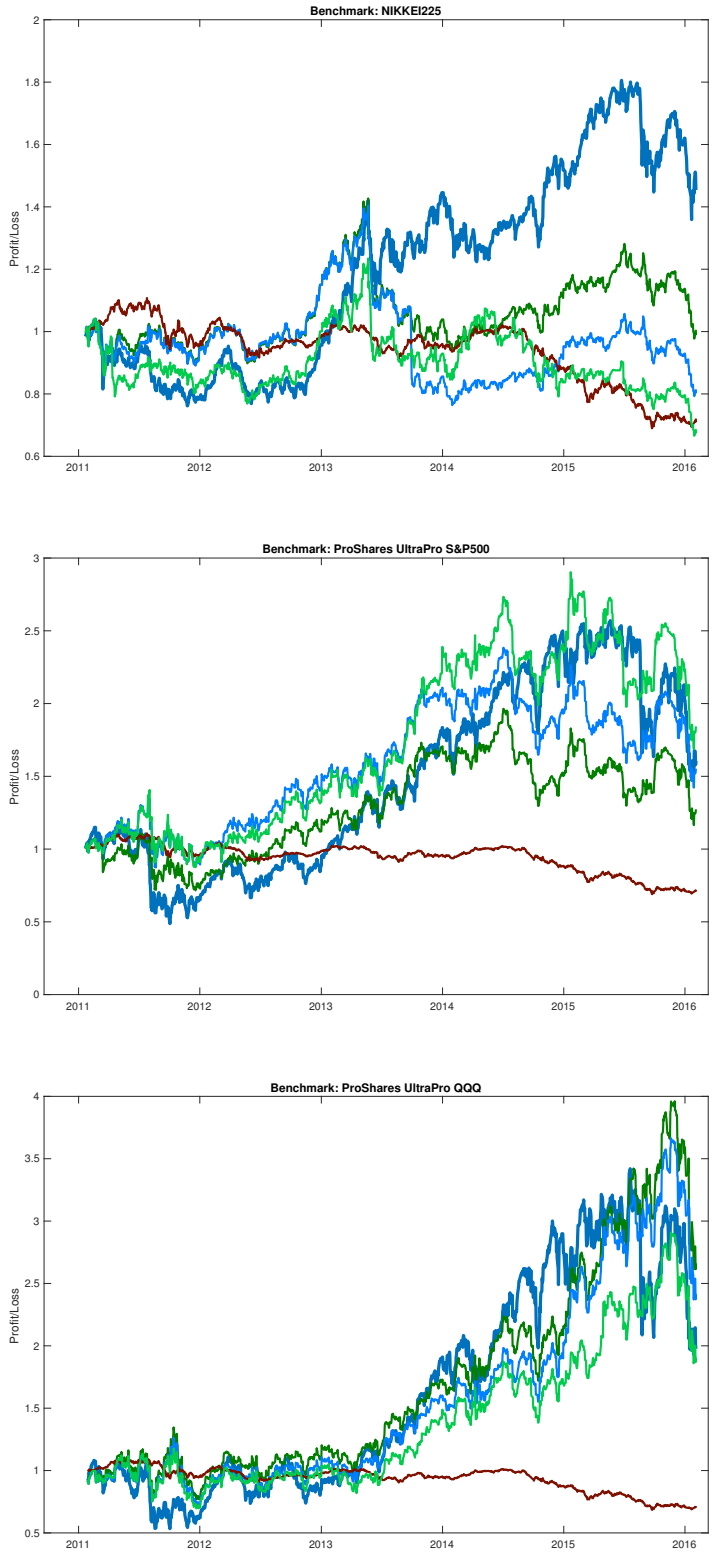

Figure 1.2: Cumulative portfolio wealth comparison: TEDAS 2, TEDAS 3, TEDAS 4, $1/p$; upper panel: NIKKEI225 B&H, middle panel: UPRO B&H, lower panel: TQQQ B&H

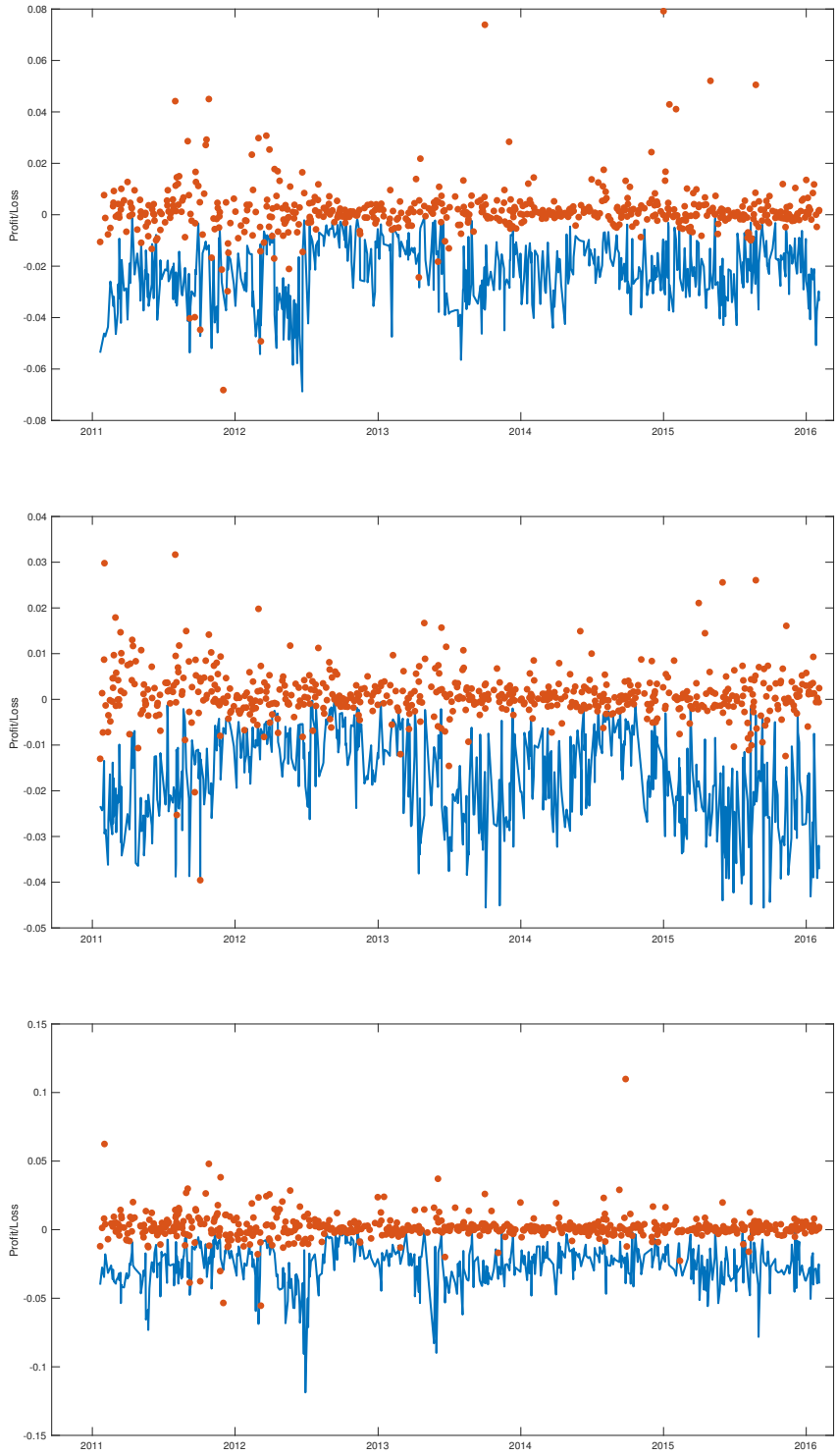

Figure 1.3: Upper panel: VaR backtest for S&P500 as the benchmark asset, Christoffersen test p-value: 0.023: middle panel: VaR backtest for FTSE100 as the benchmark asset, Christoffersen test p-value: 0.000; lower panel: VaR backtest for DAX30 as the benchmark asset: Christoffersen test p-value: 0.237; $\alpha = 0.99$ in all 3 cases

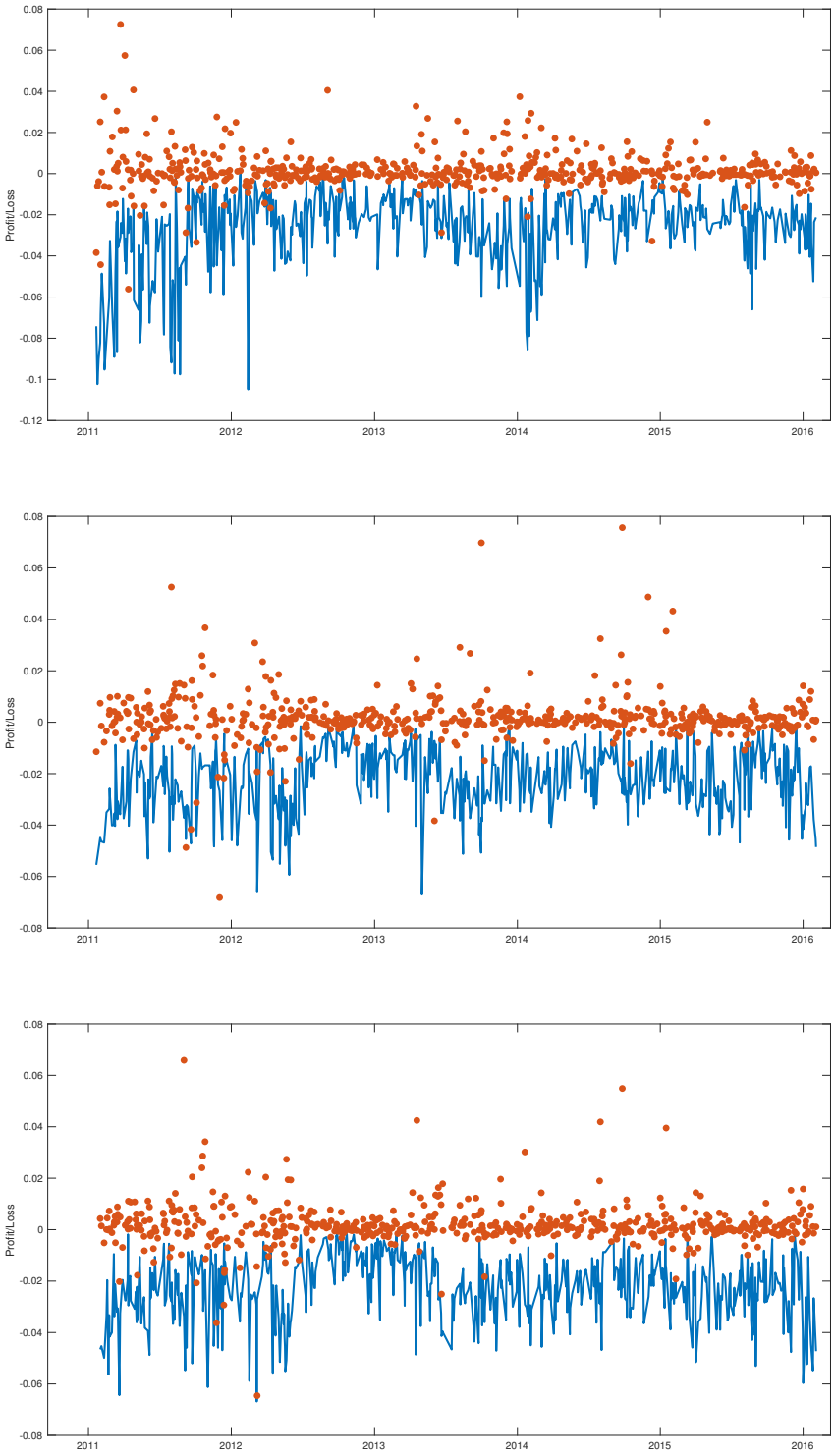

Figure 1.4: Upper panel: VaR backtest for NIKKEI225 as the benchmark asset, Christoffersen test p-value: 0.023: middle panel: VaR backtest for UPRO as the benchmark asset, Christoffersen test p-value: 0.003; lower panel: VaR backtest for TQQQ as the benchmark asset: Christoffersen test p-value: 0.199; $\alpha = 0.99$ in all 3 cases

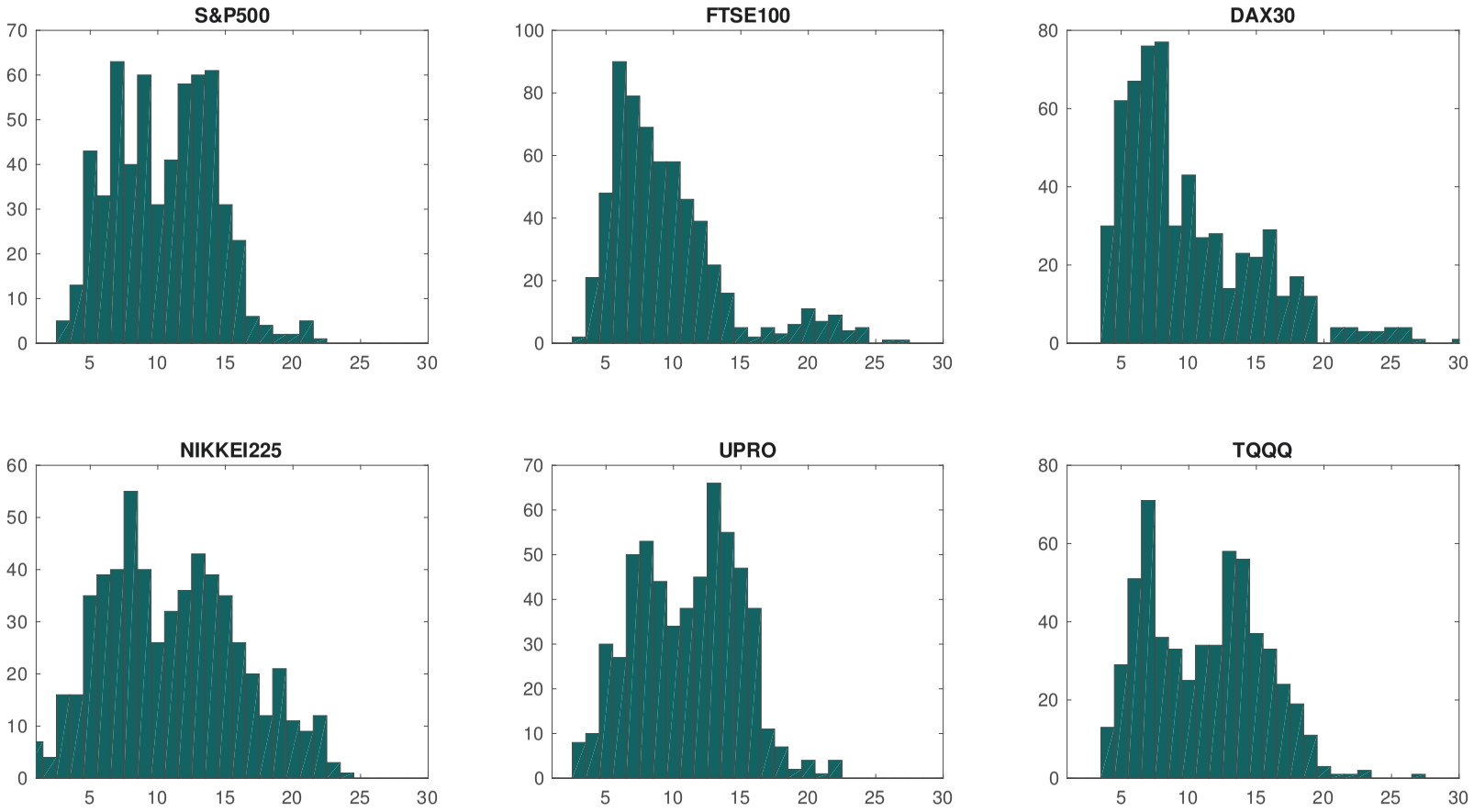

Figure 1.5: Frequency of the estimated model dimension for 6 benchmarks

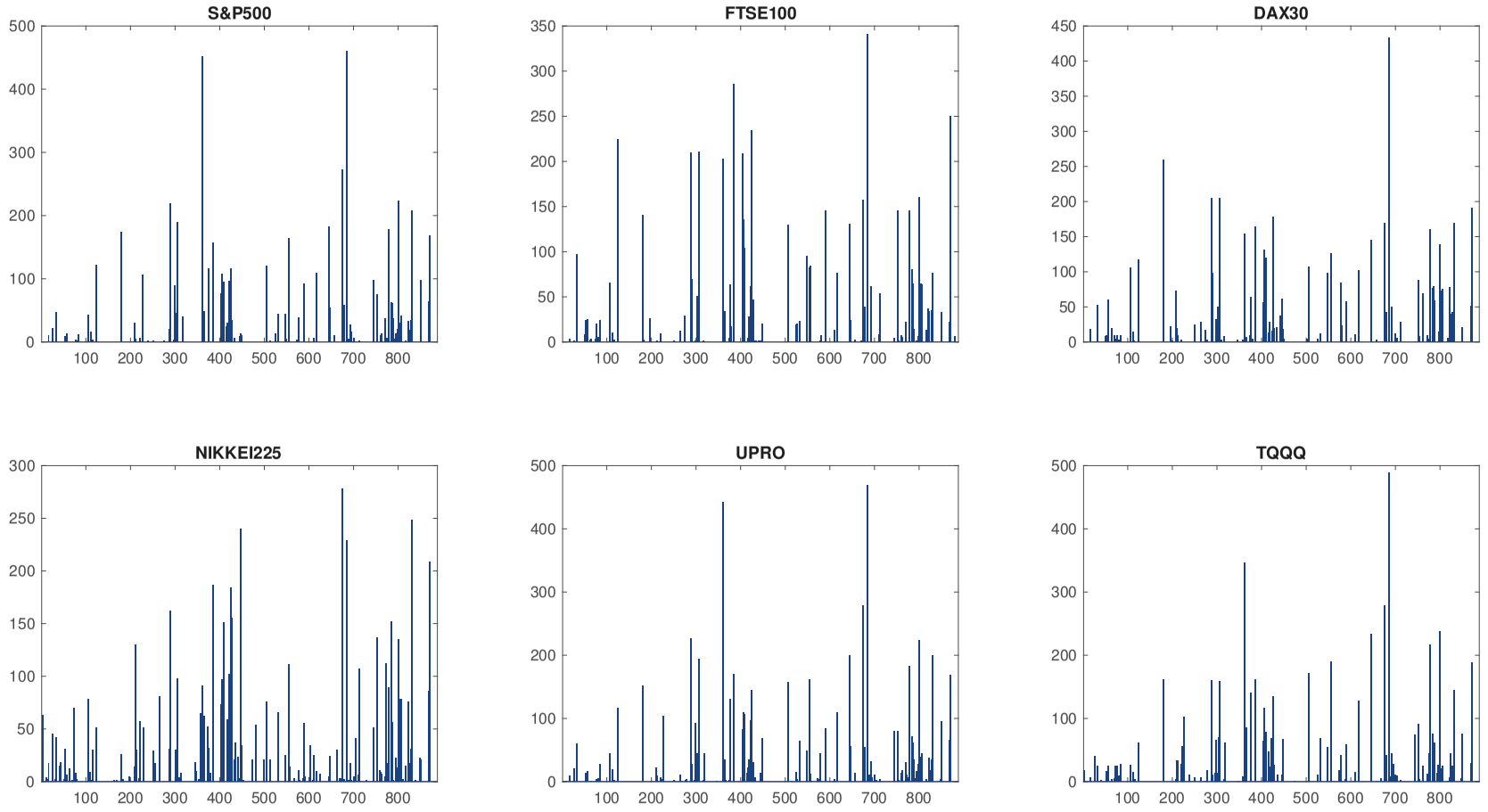

Figure 1.6: Frequency of the selected hedge funds for 6 benchmarks

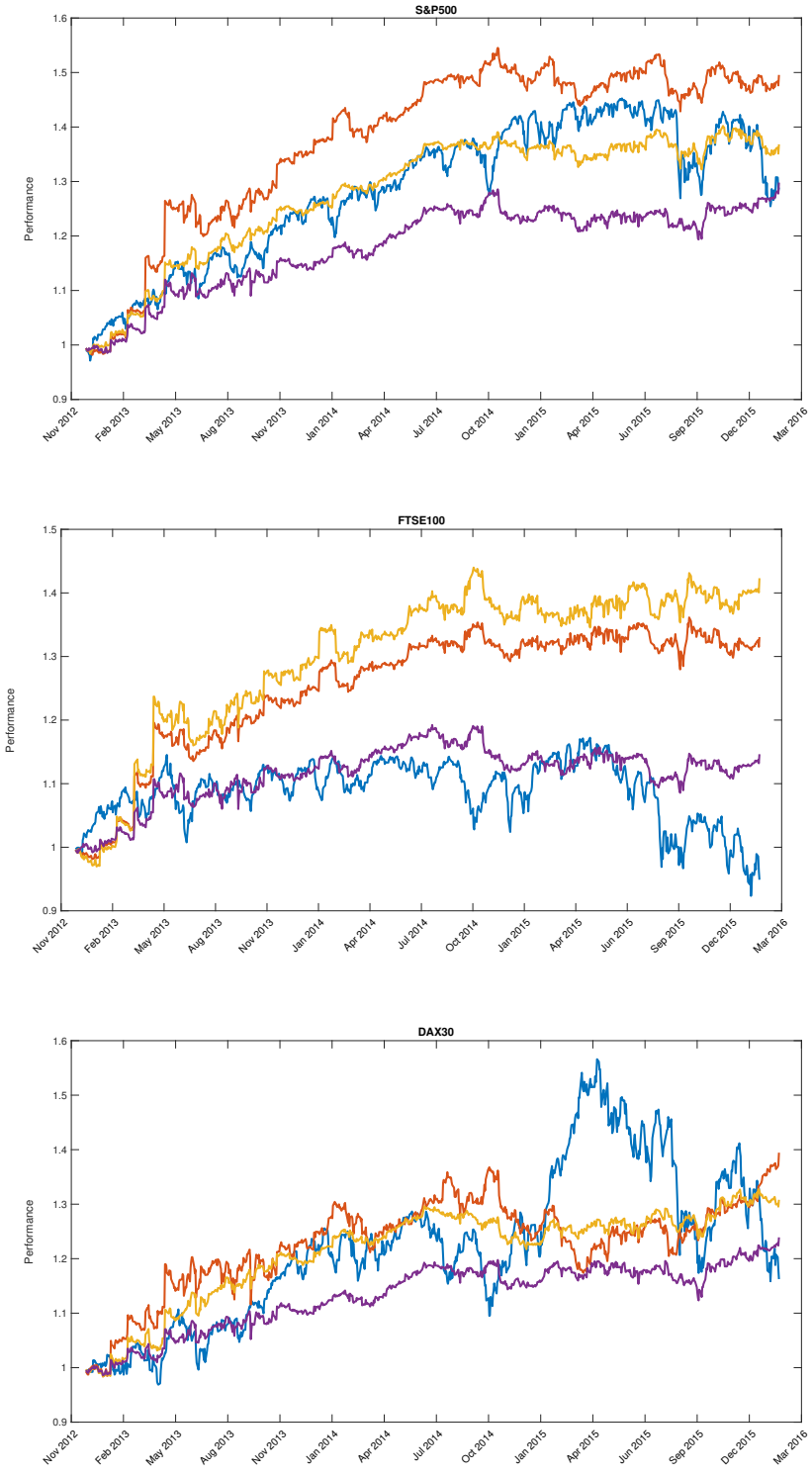

Figure 1.7: Performance of static portfolios of benchmark indices and 5 most frequently chosen hedging assets for each respective index

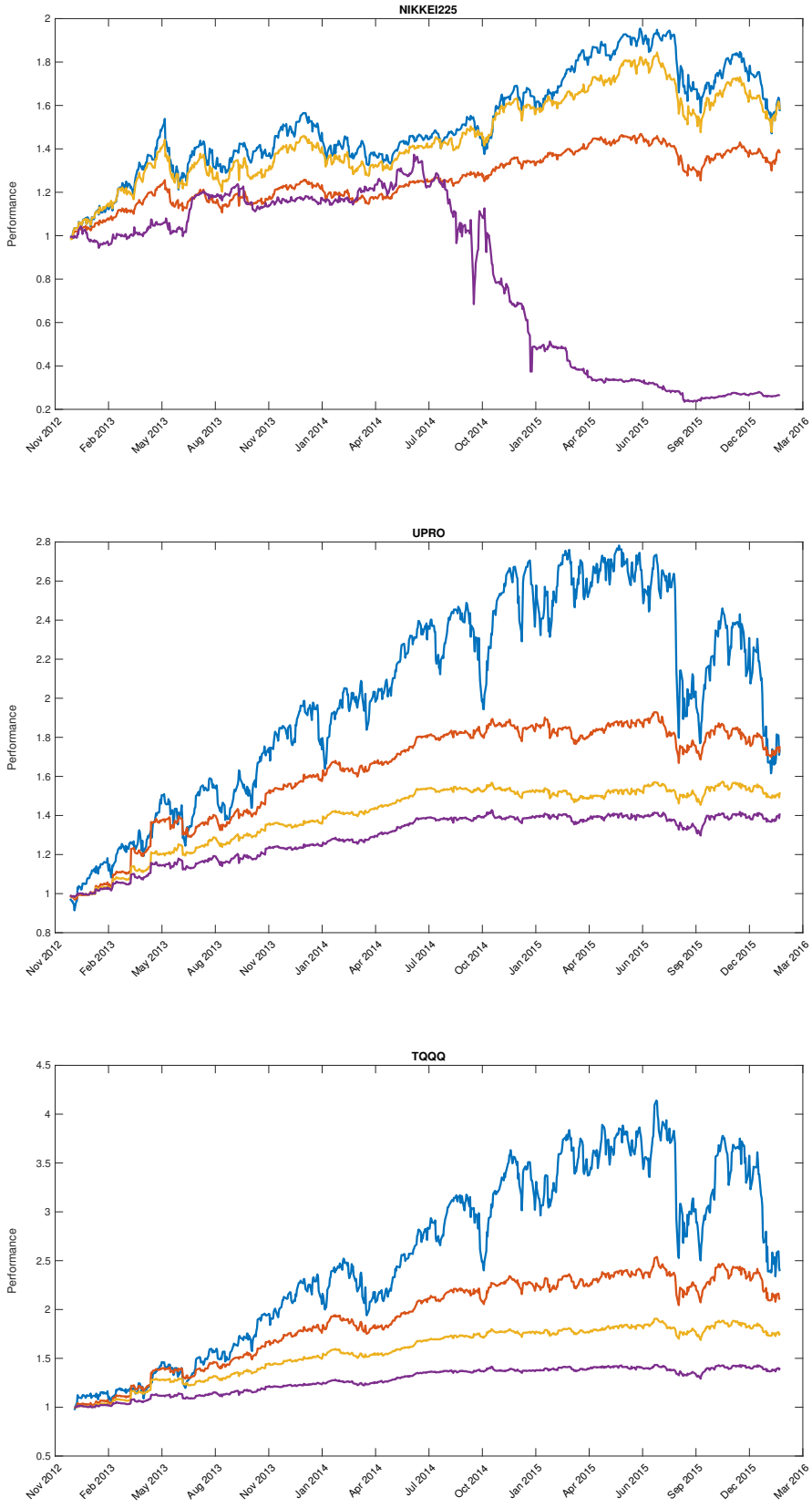

Figure 1.8: Performance of static portfolios of benchmark indices and 5 most frequently chosen hedging assets for each respective index

Chapter 2

Leveraged ETF options implied volatility paradox: a statistical study

2.1 Introduction

Exchange-traded funds (ETFs) are financial products that track indices, commodities, bonds, baskets of assets. They have become increasingly popular due to diversification benefits as well as the investor's ability to perform short-selling, buying on margin and lower expense ratios than, for instance those of mutual funds.

The trading advantages of the ETFs are enhanced through the use of gearing or leverage, when derivative products are used to generate multiple or inverse multiple returns on the underlying asset. For instance, the leveraged ETF ProShares Ultra S&P500 (SSO) with leverage ratio $\beta = +2$ is supposed to gain 2% for every 1% daily gain in the price of the S&P500 index, with a subtraction of an expense fee. An inverse leveraged ETF would invert the loss and amplify it proportionally to the ratio magnitude: the ProShares UltraShort S&P500 (SDS) with leverage ratio $\beta = -2$ would generate a 2% gain for every 1% daily loss in the price of the underlying S&P500 index. Naturally, the basic unleveraged SPDR S&P 500 ETF (SPY) returns 100% of the gain/loss of S&P 500 index, having $\beta = +1$.

Due to their growing popularity and the nature of ETF and LETF similar dynamics, recently there has been growing research on leveraged ETFs and their consistent pricing. Specifically, Leung and Sircar (2015) introduced the so-called "moneyness scaling" technique which links implied volatilities (IV) between ETF and LETF in the way that the discrepancy between the implied volatility "smile" pattern is removed. Recent empirical observations seemingly support this idea. Figure 2.1 below compares the empirical implied volatilities for the LETFs SSO, SDS, UPRO ($\beta = +3$), SPXU ($\beta = -3$) before moneyness scaling is done. The log-moneyness

LM is defined as

$$LM \stackrel{\text{def}}{=} \log\left(\frac{K}{L_t}\right), \tag{2.1}$$

where K is the strike of the LETF option and L_t the LETF price at time t. After implied volatility re-scaling according to the identity

$$\sigma^{resc.}_{LETF} \stackrel{\text{def}}{=} |\beta|^{-1} P^{-1}_{BS}(P_\beta), \tag{2.2}$$

where P_{BS} is the option Black-Scholes price and P_β is an observed market price of the LETF, there are still visible discrepancies between the implied volatilities for the SPY ETF and its leveraged counterparts. The moneyness scaling procedure yields a more coherent picture as in Figure 2.2, when the LETF and ETF implied volatilities overlap significantly better.

The question arises whether the moneyness scaling method indeed removes discrepancies consistently in time. To answer this question, a study is required to verify whether IV deviations are significant from the statistical point of view. This leads to the problem of constructing confidence intervals (or confidence bands) for the difference of IV estimators. Several studies including Cont and da Fonseca (2002), Aït-Sahalia et al. (2001) apply non- and semiparametric approaches to model implied volatilities. The use of such estimators allows to construct uniform or bootstrap confidence bands which can be used as a check for the potential existence of price discrepancies among ETF options with different leverage ratios.

2.2 Consistency study for moneyness scaling

2.2.1 Implied volatility as estimator

The moneyness scaling technique proposed by Leung and Sircar (2015) offers a "coordinate transformation" for the LETF option implied volatility and potentially reflects the increase of risk in the underlying asset (ETF). Based on the assumption that the distribution of the terminal price of the β-LETF depends on the leverage ratio β, the moneyness scaling formula includes an expectation of the β-LETF log-moneyness conditional on the terminal value of the unleveraged counterpart. For the LETF log-moneyness $LM^{(\beta)}$ (consider ETFs as LETFs with $\beta = 1$) the formula linking the log-moneyness coordinates $LM^{(\beta)}$ and $LM^{(1)}$ of the leveraged and unleveraged ETF, respectively:

$$\begin{aligned} LM^{(\beta)} &= \beta LM^{(1)} - \{r(\beta-1)+c\}T - \frac{\beta(\beta-1)}{2}\mathsf{E}^{\mathsf{Q}}\left\{\int_0^T \sigma_t^2 \mathrm{d}t \,\middle|\, \log\left(\frac{S_T}{S_0}\right) = LM^{(1)}\right\} \\ &= \beta LM^{(1)} - \{r(\beta-1)+c\}T - \frac{\beta(\beta-1)}{2}\sigma^2 T, \end{aligned} \tag{2.3}$$

where T is time-to-maturity, asset volatility σ is assumed constant.

More generally, for two LETFs with different leverage ratios β_1, β_2 the expression (2.3) takes the (approximate) form:

$$LM^{(\beta_1)} = \frac{\beta_1}{\beta_2}\left[LM^{(\beta_2)} + \{r(\beta_2 - 1) + c_2\}T + \frac{\beta_2(\beta_2 - 1)}{2}\sigma^2 T\right] - \{r(\beta_1 - 1) + c_1\}T - \frac{\beta_1(\beta_1 - 1)}{2}\sigma^2 T \quad (2.4)$$

Another popular measure for moneyness is the so-called forward moneyness which is an appropriate choice for European option data because European options can be only exercised at expiry. It is defined as follows:

$$\kappa_f \stackrel{\text{def}}{=} K/\{e^{(r-c)\tau} L_t\}, \quad (2.5)$$

where $r - c$ is the cost of carry, $\tau \stackrel{\text{def}}{=} T - t$. In terms of the forward moneyness, the moneyness scaling equation for two LETFs with different leverage ratios β_1, β_2 can be shown (see Appendix 4.4) to satisfy:

$${\kappa_f}^{(\beta_1)} = \exp\left\{-\frac{\beta_1}{2}(\beta_1 - \beta_2)\sigma^2\tau\right\}({\kappa_f}^{(\beta_2)})^{\frac{\beta_1}{\beta_2}}. \quad (2.6)$$

Many researchers, among them Fengler et al. (2007), Park et al. (2009) have studied the implied volatility as a random process in time, so that the data generating process includes some non-parametric function m:

$$Y_t = m(X_t) + \varepsilon_t, \quad t = 1, \ldots, T, \quad (2.7)$$

or can be driven by a latent factor process $\mathscr{Z}_t$:

$$Y_t = \mathscr{Z}_t^\top m(X_t) + \varepsilon_t, \quad t = 1, \ldots, T, \quad (2.8)$$

where Y_t stands for an implied volatility process, the covariates X_t can be one- or multi-dimensional. Usually X_t is assumed to contain a moneyness component such as κ_f and the time-to-maturity τ.

2.2.2 Confidence bands

The statistical properties of the estimators $\widehat{m}(X_t)$ and $\widehat{\mathscr{Z}_t^\top}\widehat{m}(X_t)$ for the models (2.7) and (2.8) have been outlined, respectively, in, e.g., Härdle (1990), Ruppert and Wand (1994) and Park et al. (2009). To study the consistency of the implied volatility difference between the ETF and the moneyness-scaled LETF case, one

needs to consider statistical differences of the corresponding estimators. Confidence band analysis may provide a first insight into the matter. An important issue about smooth confidence bands for functions is the correct probability of covering the "true" curve. One way to address it is to use the Bonferroni correction to adjust confidence levels for each pointwise confidence interval to obtain the overall confidence. On the other hand, asymptotic confidence bands generally tend to underestimate the true coverage probability, see Hall and Horowitz (2013).

An alternative approach is to use bootstrap confidence bands while the distribution of the original data is "mimicked" via a pre-specified random mechanism achieving both uniformity and better coverage. The approach of Härdle et al. (2015) proposes a uniform bootstrap bands construction for a wide class of non-parametric M and L-estimates. It is logical to use a robust M-type smoother for the estimation of (2.7) in the actual case of implied volatility, as IV data often suffer from outliers. The procedure runs as follows: consider the sample $\{X_t, Y_t\}_{t=1}^T$, where Y_t denotes the IV process, X_t is taken to be one-dimensional and includes the log-moneyness covariate LM^β.

1. Compute the estimate $\widehat{m}_h(X)$ by a local linear M-smoothing procedure (see Appendix 4.5) with some kernel function and bandwidth h chosen by, e.g., cross-validation.

2. Given $\widehat{\varepsilon}_t \stackrel{\text{def}}{=} Y_t - \widehat{m}_h(X_t)$; $\widehat{m}$ obtained in Step 1, do bootstrap resampling from $\widehat{\varepsilon}_t$, that is, for each $t = 1, \ldots, T$, generate a random variable $\varepsilon_t^* \sim \widehat{F}_{\varepsilon|X}(z)$ and a re-sample
$$Y_t^* = \widehat{m}_g(X_t) + \varepsilon_t^*, \quad t = 1, \ldots, T \tag{2.9}$$
B times (bootstrap replications) with an "oversmoothing" bandwidth $g \gg h$ such as $g = \mathcal{O}(T^{-1/9})$ to allow for a bias correction.

3. For each re-sample $\{X_t, Y_t^*\}_{t=1}^T$ compute $\widehat{m}_{h,g}^*$ using the bandwidth h and construct the random variable
$$d_b \stackrel{\text{def}}{=} \sup_{x \in B} \left[\frac{|\widehat{m}_{h,g}^*(x) - \widehat{m}_g(x)| \sqrt{\widehat{f}_X(x)} \widehat{f}_{\varepsilon|X=x_t}(\varepsilon_t^*)}{\sqrt{\widehat{\mathsf{E}}_{Y|X}\{\psi^2(\varepsilon_t^*)\}}} \right], \quad b = 1, \ldots, B, \tag{2.10}$$
where B is a finite compact support set of $\widehat{f}_X$ and $\psi(u) = \rho'(\cdot)$, see Appendix 4.5.

4. Calculate the $1 - \alpha$ quantile d_α^* of $d_1, \ldots, d_B$.

5. Construct the bootstrap uniform confidence band centered around $\widehat{m}_h(x)$:
$$\widehat{m}_h(x) \pm \left[\frac{\sqrt{\widehat{\mathsf{E}}_{Y|X}\{\psi^2(\varepsilon_t^*)\}} d_\alpha^*}{\sqrt{\widehat{f}_X(x)} \widehat{f}_{\varepsilon|X=x_t}(\varepsilon_t^*)} \right] \tag{2.11}$$

Daily data are used in the period 20141117-20151117 to construct bootstrap confidence bands for the M-smoother of implied volatility Y given forward moneyness X. For the LETFs SSO, UPRO, SDS X is transformed via moneyness scaling formula (2.6). The results are shown in Figures 2.3, 2.4 and 2.5 for time-to-maturity 0.5, 0.6 and 0.7 years, respectively. A closer look at the obtained bootstrap confidence bands' relative location might imply that the moneyness scaling procedure can remove the discrepancy between the implied volatilities of leveraged ETFs and their unleveraged counterpart with possible deviations for two positively leveraged ETFs (SSO and UPRO) for smaller moneyness values. Should arbitrage opportunities arise, they are quickly traded away, given that the markets for the presented ETFs are quite liquid. However, caution should be exercised as the joint hypothesis on the difference of two estimators with a single band may still be rejected.

The illustration of the confidence bands' overlap in Figure 2.6 shows that the bands for the SSO LETF become wider than before and are not covered by those of the unleveraged counterpart. This implies possible discrepancies not removed by the moneyness scaling procedure. Given that the moneyness scaling approximation (2.6) is correct, there are arbitrage opportunities in the market of SPY and SSO which may be exploited by traders at different times to maturity.

2.3 Moneyness scaling under stochastic volatility

2.3.1 A semi-analytical approach

In the study of moneyness scaling, one needs to estimate the following conditional expectation:

$$\mathsf{E}^{\mathsf{Q}}\left\{\int_0^T \sigma_t^2 \mathrm{d}t \,\middle|\, \log\left(\frac{S_T}{S_0}\right) = LM^{(1)}\right\}. \tag{2.12}$$

As has been seen, taking $\sigma_t = \sigma$ constant, one obtains $\sigma^2 T$. As empirical evidence shows, this is not a plausible assumption, therefore one needs to determine the measure Q for the case of random volatility under a model selected as "properly" as possible. Second, one needs to estimate the quantity

$$\int_0^t \sigma_s^2 \mathrm{d}s, \tag{2.13}$$

which is called the integrated variance or the integrated volatility and its estimation is a subject of considerable interest in financial literature.

Stochastic volatility presents a viable alternative to the constant case. One could choose among different specifications of stochastic volatility models. A general stochastic volatility model is defined through a system of stochastic differential

equations, e.g.:

$$dS_t = A(t,S_t,V_t)dt + \xi(t,S_t,V_t)dW_{S,t}, \tag{2.14}$$
$$dV_t = B(t,S_t,V_t)dt + \sigma(t,S_t,V_t)dW_{V,t}, \tag{2.15}$$

for given functions $A(\cdot)$, $B(\cdot)$, $\sigma(\cdot)$, $\xi(\cdot)$ and a two-dimensional standard Wiener process $W_t \overset{\text{def}}{=} (W_{S,t}, W_{V,t})^\top$ such that $S_0 = s_0$, $V_0 = v_0$, $t \geq 0$, $s_0, v_0 \in \mathbb{R}$. Some special cases are models such as Heston (1993), Hull and White (1987), Schöbel and Zhu (1999) specifications. An example of a stochastic volatility system which is built of more than 2 equations, is given in Leung and Sircar (2015). However, simpler models tend to generate semi-closed-form solutions for the distributions of the log-returns $x_t \overset{\text{def}}{=} \log(S_t/S_{t-1})$. For the Heston model, a solution was proposed by Dragulescu and Yakovenko (2002).

The Heston model with risk-neutral dynamics under a risk-neutral measure Q and zero volatility risk premium is described by a two-dimensional system of stochastic differential equations

$$dS_t = (r-c-0.5)S_t dt + \sqrt{V_t}S_t dW_{S,t}^{\mathsf{Q}}, \tag{2.16}$$
$$dV_t = \kappa(\theta - V_t)dt + \sigma dW_{V,t}^{\mathsf{Q}}, \tag{2.17}$$

where $r-c$ are costs of carry on S_t, θ is the long-run variance level, κ is the rate of reversion to θ, σ is the "volatility of the volatility" parameter which determines the variance of V_t; $W_{S,t}$, $W_{V,t}$ are correlated with parameter ρ.

In Figures 2.7, 2.8 the densities of daily returns of the SCO LETF (ProShares UltraShort Bloomberg Crude Oil) are plotted against the densities implied by the geometric Brownian motion. It can be noticed that the tails of the Heston-implied densities are exponential and heavier than those of the normal distribution with the dispersion parameter equal to the long-term variance θ.

The solution to (2.12) can be computed in terms of inverse Fourier transforms of characteristic functions for particular stochastic volatility models.

Proposition 2.3.1.1. Under the Heston model given by (2.16)-(2.17), assuming zero costs of carry, (2.12) equals:

$$\left[\int_0^T \frac{\begin{array}{r}\int_0^\infty \Re\left\{\frac{S_0^{-i\phi}e^{-i\phi LM^{(1)}}}{i\phi} f_1(\phi)\left(\frac{\partial C_1}{\partial t}+\frac{\partial D_1}{\partial t}V_0\right)\right. \\ \left. -f_2(\phi)\left(\frac{\partial C_2}{\partial t}+\frac{\partial D_2}{\partial t}V_0\right)\right\}\mathrm{d}\phi\end{array}}{\begin{array}{r}\int_0^\infty \Re\left\{S_0^{-(i\phi+2)}e^{-LM^{(1)}(i\phi+2)}\right. \\ \left. \times f_1(\phi)+S_0 e^{LM^{(1)}}(i\phi-1)f_2(\phi)\right\}\mathrm{d}\phi\end{array}} dt\right] \times \left(\frac{2}{S_0^2 e^{2LM^{(1)}}}\right), \tag{2.18}$$

where $\Re$ denotes the real part of a number,

$$
\begin{aligned}
f_{1,2}(\phi) &= \exp(C_{1,2} + D_{1,2}V_0 + i\phi S_t), \\
C_{1,2} &= ri\phi\tau + \frac{\kappa\theta}{\sigma^2}\left\{(b_{1,2} - \rho\sigma i\phi - d_{1,2})\tau - 2\log\left(\frac{1 - c_{1,2}e^{-d_{1,2}\tau}}{1 - c_{1,2}}\right)\right\}, \\
D_{1,2} &= \frac{(b_{1,2} - \rho\sigma i\phi - d_{1,2})}{\sigma^2}\left(\frac{1 - e^{-d_{1,2}\tau}}{1 - c_{1,2}e^{-d_{1,2}\tau}}\right), \\
\frac{\partial C_{1,2}}{\partial t} &= ri\phi + \frac{\kappa\theta}{\sigma^2}\left\{(b_{1,2} - \rho\sigma i\phi + d_{1,2}) + 2\left(\frac{g_{1,2}d_{1,2}e^{d_{1,2}t}}{1 - g_{1,2}e^{d_{1,2}t}}\right)\right\},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial D_{1,2}}{\partial t} &= \frac{(b_{1,2} - \rho\sigma i\phi + d_{1,2})}{\sigma^2}\left(\frac{d_{1,2}e^{d_{1,2}t}(g_{1,2}e^{d_{1,2}t)-1} + (1 - e^{d_{1,2}t})g_{1,2}d_{1,2}e^{d_{1,2}t}}{(1 - g_{1,2}e^{d_{1,2}t})^2}\right), \\
d_{1,2} &= \sqrt{(\rho\sigma i\phi - b_{1,2})^2 - \sigma^2(2u_{1,2}i\phi - \phi^2)}, \;\; c_{1,2} = \frac{b_{1,2} - \rho\sigma i\phi - d_{1,2}}{b_{1,2} - \rho\sigma i\phi + d_{1,2}} \\
b_1 &= \kappa - \rho\sigma, \;\; b_2 = \kappa, \;\; u_1 = 0.5, \;\; u_2 = -0.5.
\end{aligned}
$$

Proof. See Appendix 4.6. □

2.3.2 A Monte-Carlo approach

Alternatively, the conditional expectation in (2.12) can be computed using Monte-Carlo simulations. The simulations are performed using the Heston model and the calibrated parameters obtained minimizing the squared difference between theoretical Heston prices $C^{\Theta}(K,\tau)$ obtained from the model and observed market prices $C^{M}(K,\tau)$,

$$
\min_{\Theta\in\mathbb{R}^5} \sum_{i=1}^{N} \left(C_i^{\Theta}(K_i,\tau_i) - C_i^{M}(K_i,\tau_i)\right) \tag{2.19}
$$

where $\Theta \overset{\text{def}}{=} (\kappa,\theta,\sigma,v_0,\rho)$ Heston parameters, N number of options used for calibration, K strikes and τ times-to-maturity. Theoretical prices $C^{\Theta}(K,\tau)$ are obtained via numeric integration of the Heston characteristic function. The estimation results on day 20160205 for the SCO LETF yield $\widehat{\kappa} = 2.50$, $\widehat{\theta} = 0.53$, $\widehat{\sigma} = 1.32$, $\widehat{v_0} = 1.41$, $\widehat{\rho} = 0.48$.

The Monte-Carlo algorithm is motivated by van der Stoep et al. (2014) and can be formulated as follows:

1. Generate N pairs of observations (s_i, v_i), $i = 1,\ldots,N$.

2. Order the realizations s_i: $s_1 \leq x_2 \leq \ldots \leq s_N$.

3. Determine the boundaries of M bins $(l_k, l_{k+1}]$, $k = 1, \ldots, M$ on an equidistant grid of values $S^* \stackrel{\text{def}}{=} S_0 e^{LM^{(1)}}$

4. For the kth bin approximate the conditional expectation (2.4) by

$$\mathsf{E}^{\mathsf{Q}}\left(\int_0^T \sigma_s^2 \mathrm{d}s \,|S_T \in (l_k, l_{k+1}]\right) \approx \frac{h}{NQ(k)} \sum_{i=1}^{H} \sum_{j \in \mathscr{J}_k} V_{ij}, \tag{2.20}$$

where h is the discretization step for V_t, $\mathscr{J}_k$ the set of numbers j, for which the observations S_T are in the kth bin and $Q(k)$ is the probability of S_T being in the kth bin.

The results of the simulation are presented in Figure 2.9. Polynomial smoothing is applied to produce the smoothed version of SCO LETF integrated volatility. The generated expected integrated variance has the form of a "smile" which confirms the intuition behind using average square implied volatility in the case of constant-volatility moneyness scaling approach.

2.4 Dynamic option trading strategy

2.4.1 The dynamic semiparametric factor model setup

Model description

A generalized version of the model in (2.7) represented by (2.8) assumes the implied volatility Y_t to be a stochastic process driven by a latent stochastic factor process $\mathscr{Z}_t$ contaminated by noise ε_t. To be more specific, define $\mathscr{J} \stackrel{\text{def}}{=} [\kappa_{min}, \kappa_{max}] \times [\tau_{min}, \tau_{max}]$, $Y_{t,j}$ implied volatility, $t = 1, \ldots, T$ time index, $j = 1, \ldots, J_t$ option intraday numbering on day t, $X_{t,j} \stackrel{\text{def}}{=} (\kappa_{t,j}, \tau_{t,j})^\top$, $\kappa_{t,j}$, $\tau_{t,j}$ are, respectively, a moneyness measure (log-, forward, etc.) and time-to-maturity at time point t for option j. Then the *dynamic semiparametric factor model* is defined as follows: assume

$$Y_{t,j} = \mathscr{Z}_t^\top m(X_{t,j}) + \varepsilon_{t,j}, \tag{2.21}$$

where $\mathscr{Z}_t = (1, Z_t^\top)$, $Z_t = (Z_{t,1}, \ldots, Z_{t,L})^\top$ unobservable L-dimensional stochastic process, $m = (m_0, \ldots, m_L)^\top$, real-valued functions; m_l, $l = 1, \ldots, L+1$ are defined on a subset of $\mathbb{R}^d$. The full description of the model is given in Park et al. (2009). One can estimate:

$$\widehat{Y}_t = \widehat{\mathscr{Z}_t^\top} \widehat{m}(X_t) \tag{2.22}$$

$$= \widehat{\mathscr{Z}_t^\top} \widehat{\mathscr{A}} \psi(X_t), \tag{2.23}$$

with $\psi(X_t) \stackrel{\text{def}}{=} \{\psi_1(X_t), \ldots, \psi_K(X_t)\}^\top$ being a space basis such as a tensor B-spline basis, $\mathscr{A}$ is the $(L+1) \times K$ coefficient matrix. In this case K denotes the number of tensor B-spline sites: let $(s_u)_{u=1}^U$, $(s_v)_{v=1}^V$ be the B-spline sites for moneyness and time-to-maturity coordinates, respectively, then $K = U \cdot V$. Given some spline orders n_κ and n_τ for both coordinates and sets of knots $(t_i^\kappa)_{i=1}^M$, $(t_j^\tau)_{j=1}^N$, one of the Schoenberg-Whitney conditions requires that $U = M - n_\kappa$, $V = N - n_\tau$, see De Boor (2001). The usage of the parameter K is roughly analogous to the bandwidth choice in Fengler et al. (2003) and Fengler et al. (2007); however the results of Park et al. (2009) demonstrate insensitivity of DSFM estimation results to the choice of K, n.

The estimates for the IV surfaces $\widehat{m}_l$ are re-calculated on a fine 2-dimensional grid of tensor B-spline sites: the estimated coefficient matrix $\widehat{\mathscr{A}}$ is reshaped into a $U \times V \times L+1$ array of $L+1$ matrices $\widehat{A}$ of dimension $U \times V$. Factor functions m_l can then be estimated as follows:

$$\widehat{m}_{l;i,j} = \sum_i^U \sum_j^V \widehat{A}_{l;i,j} \psi_{i,k_\kappa}(\kappa_i) \psi_{j,k_\tau}(\tau_j), \tag{2.24}$$

where k_κ, k_τ are knot sequences for the moneyness and time-to-maturity coordinates, respectively.

The estimated factor functions $\widehat{m}_l$ together with stochastic factor loadings $\widehat{\mathscr{Z}_t}$ are combined into the dynamic estimator of the implied volatility surface:

$$\widehat{IV}_{t;i,j} = \widehat{m}_{0;i,j} + \sum_{l=1}^L \widehat{\mathscr{Z}_{l,t}} \widehat{m}_{l;i,j}, \tag{2.25}$$

can be modeled as a vector autoregressive process. It should be noted that $\widehat{m}_l$ and $\widehat{\mathscr{Z}_{l,t}}$ are not uniquely defined, so an orthonormalization procedure must be applied.

An indication of possible mispricing of LETF options allows to test an investment strategy based on the comparison of the theoretical price obtained from the moneyness scaling correction as well as the application of the DSFM model and the market price. Such a strategy would mainly exploit the two essential elements of information from these two approaches. The first element is obtaining an indication of arbitrage opportunities resulting from the mismatch between ETF and LETF IVs. The moneyness scaling approach allows to estimate LETF IV using richer unleveraged ETF data which also would make the DSFM IV estimator more consistent. The second element is implied volatility forecasting. The DSFM model allows to forecast a whole IV surface via the dynamics of stochastic factor loadings $\mathscr{Z}_t$.

Empirical results

For the model's empirical testing, the data on SPY and SSO (L)ETF European call options are used in the period 20140920-20150630. The data summary statistics are outlined in Table 2.1 below.

To avoid computational problems, the estimation space $[\kappa_{min}, \kappa_{max}] \times [\tau_{min}, \tau_{max}]$ for covariates $X_t = (\kappa_t, \tau_t)^\top$, which covers in (forward) moneyness $\kappa \in [0.3, 1.5]$ and in time-to-maturity $\tau \in [0.3, 1.0]$, is re-scaled (via marginal distribution functions) to $[0,1]^2$. The model is estimated using numerical methods, see Park et al. (2009). The number of the dynamic functions has to be chosen in advance. One should also notice that for m_l to be chosen as eigenfunctions of the covariance operator $K(u,v) \stackrel{\text{def}}{=} \mathsf{Cov}\{Y(u), Y(v)\}$ in an L-dimensional approximating linear space, where Y is understood to be the random IV surface, they should be properly normalized, such that $\|m_l(\cdot)\| = 1$ and $\langle m_l, m_k \rangle = 0$ for $l \neq k$.

The choice of L can be based on the explained variance by factors:

$$EV(L) \stackrel{\text{def}}{=} 1 - \frac{\sum_{t=1}^{T} \sum_{j=1}^{J_t} \left\{ Y_{t,j} - \sum_{l=0}^{L} \widehat{Z}_{t,l} \widehat{m}_l(X_{t,j}) \right\}^2}{\sum_{t=1}^{T} \sum_{j=1}^{J_t} (Y_{t,j} - \overline{Y})^2} \tag{2.26}$$

The model's goodness-of-fit is evaluated by the root mean squared error (RMSE) criterion:

$$RMSE \stackrel{\text{def}}{=} \sqrt{\frac{1}{\sum_t J_t} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \left\{ Y_{t,j} - \sum_{l=0}^{L} \widehat{Z}_{t,l} \widehat{m}_l(X_{t,j}) \right\}^2} \tag{2.27}$$

The $EV(L)$ and $RMSE$ criteria are displayed in Table 2.2 below. The model order $L = 3$ is chosen for model estimation. The data for the SPY ETF option are used with parameters $n_\kappa, n_\tau = 3$; $M = 9$, $N = 7$, so that $U = 6$, $V = 4$, $K = 6 \cdot 4 = 24$. The estimates for the factor functions m_l according to (2.24) are plotted in Figure 2.11 below. Furthermore, the statistical properties of the stochastic factor loadings $\widehat{\mathscr{Z}_t}$ estimated by the model can be studied. Figure 2.10 shows the dynamics of $\widehat{\mathscr{Z}_t}$ in time. Significant "spikes" in the beginning of the period correspond to the period of relatively large values of the VIX index. The first of them happen before the largest increase in the VIX values in the given period, therefore the model has predictive value with respect to market instability dynamics.

Theoretical and simulation results in Park et al. (2009) justify using vector autoregression (VAR) analysis to model $\widehat{\mathscr{Z}_t}$. To select a VAR model, the Schwarz, the Hannan-Quinn and the Akaike criteria have been computed, as shown in Table 2.3. All three criteria select the VAR(1) model. Furthermore, the roots of the characteristic polynomial all lie inside the unit circle, which shows that the specified model

is stationary. Portmanteau and Breusch-Godfrey LM test results with 12 lags for the autocorrelations of the error term fail to reject residual autocorrelation at 10% significance level.

The estimates of factor functions, stochastic loadings $\widehat{m}_l$, $\widehat{\mathscr{Z}_t}$ together determine the dynamics of implied volatility surfaces $\widehat{IV}_t$, as in (2.25). As an illustration, both the observed IV "strings" and the fitted by DSFM IV surface are displayed in Figure 2.12.

The degenerate nature of implied volatility data is reflected by the fact that empirical observations do not cover estimation grids at given time points. This is due to the fact that contracts at certain maturities or strikes are not always traded. The DSFM fitting procedure introduces basis functions which approximate a high-dimensional space and depend on time. This allows to account for all information in the dataset simultaneously in one minimization procedure which runs over all $\widehat{m}_l$ and $\widehat{\mathscr{Z}_t}$ and avoid bias problems which would inevitably occur if some kernel smoothing procedure such as Nadaraya-Watson were applied for this type of degenerate data.

2.4.2 The strategy

Description

Ability to forecast the whole surface of implied volatility can be used in combination with the moneyness scaling technique to exploit potential discrepancies in ETF and LETF option prices or implied volatilities to build an arbitrage trading strategy. A suitable strategy would be the so-called "trade-with-the-smile/skew" strategy adapted for the special case of ETF-LETF option IV discrepancy. It would use the ETF option data to estimate the model (theoretical) smile of the leveraged counterpart and the information from the IV surface forecast to recognize the future (one-period-ahead) possible IV discrepancy.

Going back to the results in Section 2.2.2, one can come to a conclusion that there's a certain discrepancy between SPY and SSO option implied volatilities from the statistical point of view, so these two options are considered in the strategy setup. The strategy can be outlined as follows: choose a moving window width w; then for each $t = w, \dots, T$, T is the final time point in the sample do the following:

1. Given two leverage ratios $\beta_2 = 1$, β_1, re-scale the log-moneyness coordinate LM^{β_2} according to the moneyness scaling formula (2.4) to obtain $\widehat{LM}^{\beta_1}$. This will be the "model" moneyness coordinate for DSFM estimation.

2. Estimate the DSFM model (2.21) on the space $[\widehat{LM}^{\beta_1}_{min}, \widehat{LM}^{\beta_1}_{max}] \times [\tau^{SPY}_{min}, \tau^{SPY}_{max}]$

(re-scaled to $[0,1]^2$, as suggested above). This will yield the IV surface estimates $\widehat{IV}_1, \ldots, \widehat{IV}_t$.

3. Forecast the IV surface estimate $\widehat{IV}_{t+1}$ using the VAR structure of the estimated stochastic loadings $\widehat{\mathscr{Z}_t}$ and factor functions $\widehat{m}_l$.

4. Choose a specific IV "string" for some time-to-maturity τ^* at time point t using SSO option data and calculate the marginally transformed value $\widehat{LM}^{\beta_1}_{\tau^*}$ of the *true* SSO log-moneyness LM^{β_1} using the marginal distribution of $\widehat{LM}^{\beta_1}$.

5. Using $\widehat{LM}^{\beta_1}_{\tau^*}$, τ^* and $\widehat{IV}_{t+1}$, interpolate the "theoretical" IV $\widehat{IV}_{t+1}$ over the marginally re-scaled $[\widehat{LM}^{\beta_1}_{min}, \widehat{LM}^{\beta_1}_{max}] \times [\tau^*, \tau^*]$ to obtain "theoretical" values $\widehat{IV}_{t+1;LM^{\beta_1}_{\tau^*},\tau^*}$.

6. Compare "theoretical" values $\widehat{IV}_{t+1;LM^{\beta_1}_{\tau^*},\tau^*}$ with "true" $IV_{t;LM^{\beta_1}_{\tau^*},\tau^*}$ and construct a delta-hedged option portfolio:

   - if $\widehat{IV}_{t+1;LM^{\beta_1}_{\tau^*},\tau^*} > IV_{t;LM^{\beta_1}_{\tau^*},\tau^*}$ for the whole $LM^{\beta_1}_{\tau^*}$, then buy (long) options corresponding to the largest difference $\widehat{IV}_{t+1;LM^{\beta_1}_{\tau^*},\tau^*} - IV_{t;LM^{\beta_1}_{\tau^*},\tau^*}$
   - if $\widehat{IV}_{t+1;LM^{\beta_1}_{\tau^*},\tau^*} < IV_{t;LM^{\beta_1}_{\tau^*},\tau^*}$ for the whole $LM^{\beta_1}_{\tau^*}$, then sell (short) options corresponding to the largest difference $IV_{t;LM^{\beta_1}_{\tau^*},\tau^*} - \widehat{IV}_{t+1;LM^{\beta_1}_{\tau^*},\tau^*}$
   - if $\widehat{IV}_{t+1;LM^{\beta_1}_{\tau^*},\tau^*}$ and $IV_{t;LM^{\beta_1}_{\tau^*},\tau^*}$ intersect, then buy (long) an option with the absolute largest negative deviation from the "theoretical" IV (IV expected to fall) and sell (short) an option with the smallest positive deviation from the "theoretical" IV (IV expected to increase). In all three cases use the underlying SSO LETF asset to make the whole portfolio delta-neutral.

7. At time point $t+1$, terminate the portfolio, calculate profit/loss and repeat until time T.

The strategy described above aims to exploit the information from the discrepancies between the forecast "theoretical" (model) SSO LETF implied volatilities and the historical ("true") ones. It protects the portfolio against unfavorable moves in the underlying asset L_t through delta-hedging and aims to gain from forecast moves in another option risk factor, the implied volatility via its explicit estimation and forecasting. The basic strategy presented here can be extended in several ways: further, including higher-order, option price sensitivities may be accounted for, such as gamma, theta hedging or charm-adjusted delta hedging. The amounts of bought and sold options can also be adjusted according to investor risk and return profiles and preferences.

Empirical application

Steps 2 and 3 of the dynamic strategy described above involve estimation out-of-sample forecasting of the IV surface $\widehat{IV}_{t+1}$ using the model estimates. The model parameters are taken to be the same as in Section 2.4.1. The rolling window width is 100 and the forecasting horizon is 1 day ahead. The prediction quality at time point $t+1$ is measured by the root mean squared prediction error given by

$$RMSPE \stackrel{\text{def}}{=} \sqrt{\frac{1}{J_{t+1}} \sum_{j=1}^{J_{t+1}} \left\{ Y_{t+1,j} - \sum_{l=0}^{L} \widehat{Z}_{t+1,l} \widehat{m}_l(X_{t+1,j}) \right\}^2} \tag{2.28}$$

The starting point of rolling-window estimation of the strategy is 20150415. The plot in Figure 2.13 below shows the RMSPE measure in time for three different model orders: $L = 2,3,4$. The average RMSPEs for $L = 2,3,4$ are, respectively, 0.095, 0.096 and 0.099; they decrease slightly as the order increases which reflects a well-known finding that more parsimonious models perform better in forecasting, see Zellner et al. (2002).

The dynamic strategy performance in the period 20150415-20150701 is displayed in Figure 2.14. Out of 55 investment periods, in 13 periods long portfolios were constructed, the remaining 42 periods net short positions were taken. The strategy is a self-financing strategy: no exogenous money infusions are done in its whole course. Furthermore, the potential of this strategy is even higher than displayed because only a fraction of already accumulated total proceeds was invested continuously following the simple setup in 2.4.2, where only two options were included into the portfolio each time. The presented strategy correctly guessed the direction of SSO LETF IV moves 82% of times.

2.5 Conclusions

In this chapter, the statistical properties of the moneyness scaling transformation by Leung and Sircar (2015) are studied. This transformation adjusts the moneyness coordinate of the implied volatility smile in an attempt to remove the discrepancy between the IV smiles for levered and unlevered ETF options. Bootstrap uniform confidence bands are constructed which indicate that in a statistical sense there remains a possibility that the implied volatility smiles are still not the same, even after moneyness scaling has been performed. This presents possible arbitrage opportunities on the (L)ETF market which can be exploited by traders.

A stochastic volatility approach is proposed which aims to improve the accuracy of the moneyness scaling method by explicit estimation of the conditional expectation

of integrated stochastic volatility. Two approaches to implement this estimate are presented: via semi-analytic calculation based on Fourier transforms and by means of a Monte-Carlo method.

A dynamic "trade-with-the-smile" strategy based on a dynamic semiparametric factor model is presented. This strategy utilizes the dynamic structure of implied volatility surface allowing out-of-sample forecasting and information on unleveraged ETF options to construct theoretical one-step-ahead implied volatility surfaces. The proposed strategy has the potential to generate significant trading gains due to simultaneous use of the information from the discrepancies between the forecast "theoretical" (model) SSO LETF implied volatilities and the historical ("true") ones. It protects the portfolio against unfavorable moves in the underlying asset through delta-hedging and aims to gain from forecast moves in another option risk factor, the implied volatility via its explicit estimation and forecasting via an advanced statistical model.

		Min.	Max.	Mean	Stdd.	Skewn.	Kurt.
SPY	TTM	0.26	1.05	0.76	0.19	-0.54	2.76
	Moneyness	0.05	1.43	0.48	0.17	-0.34	3.15
	IV	0.25	1.55	0.46	0.23	1.94	7.17
SSO	TTM	0.21	1.04	0.63	0.25	0.01	1.76
	Moneyness	0.18	1.69	0.63	0.29	0.92	3.61
	IV	0.25	1.34	0.41	0.11	1.91	10.81

Table 2.1: Summary statistics on SPY, SSO (L)ETF options from 20140920 to 20150630 (in total $\sum_t J_t = 9828,7619$ datapoints, respectively). Source: Datastream

Criterion	$L=2$	$L=3$	$L=4$	$L=5$
$EV(L)$	0.915	0.921	0.925	0.930
$RMSE$	0.090	0.088	0.087	0.082

Table 2.2: Explained variance and RMSE criteria for different model order sizes

Model order n	AIC(n)	HQ(n)	SC(n)
1	-4.20^*	-4.10^*	-3.96^*
2	-4.13	-3.96	-3.72
3	-4.07	-3.83	-3.48
4	-4.03	-3.72	-3.27
5	-3.97	-3.59	-3.03

Table 2.3: The VAR model selection criteria. The smallest value is marked by an asterisk

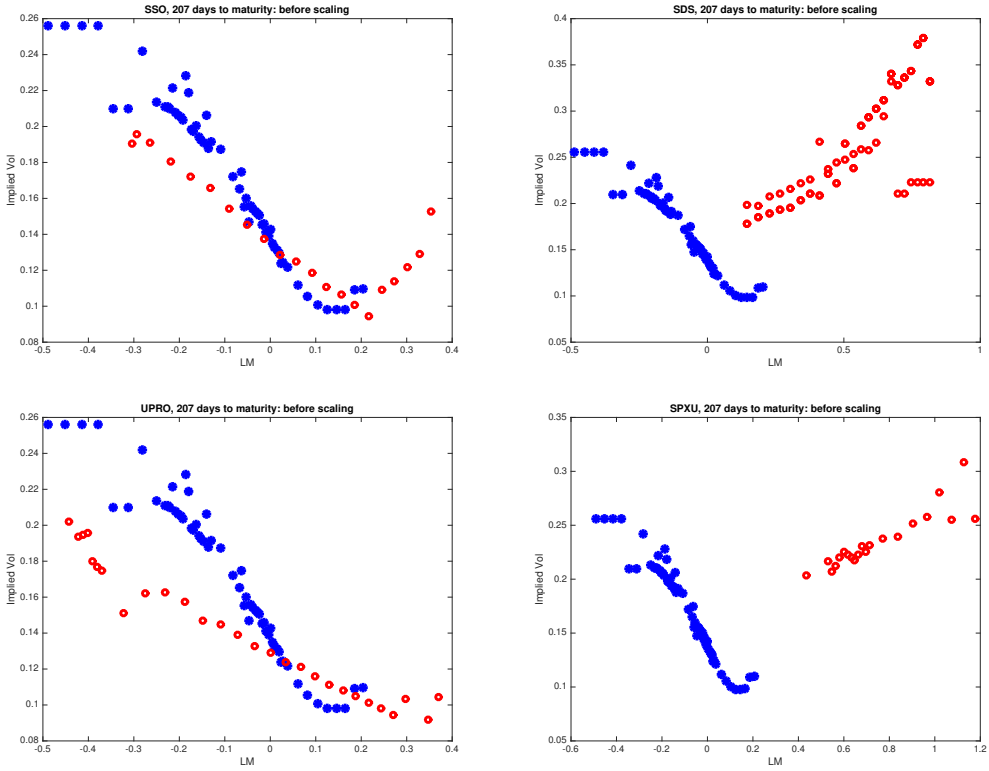

Figure 2.1: SPY (blue) and LETFs (red) implied volatilities before scaling on June 23, 2015 with 207 days to maturity, plotted against their log-moneyness

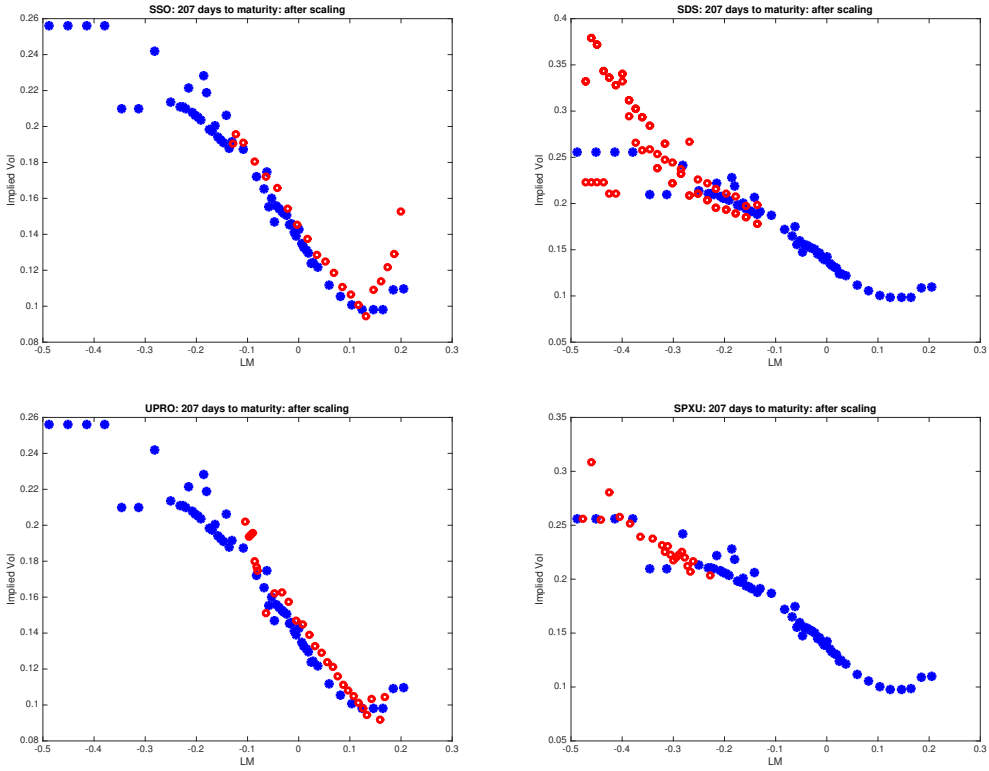

Figure 2.2: SPY (blue) and LETFs (red) implied volatilities after moneyness scaling on June 23, 2015 with 207 days to maturity, plotted against their log-moneyness

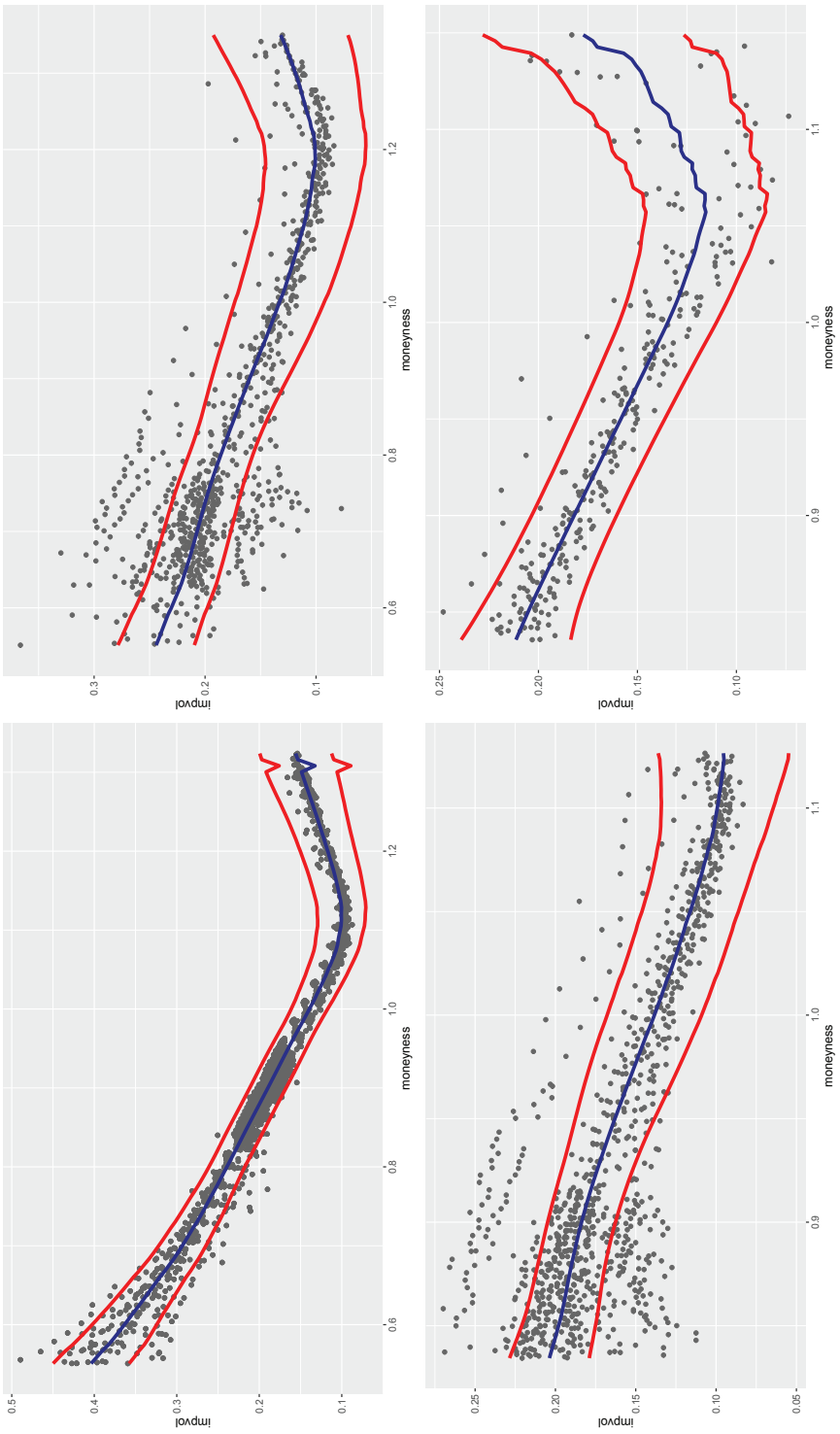

Figure 2.3: Fitted implied volatility and bootstrap uniform confidence bands for 4 (L)ETFs on S&P500; τ: 0.5 years

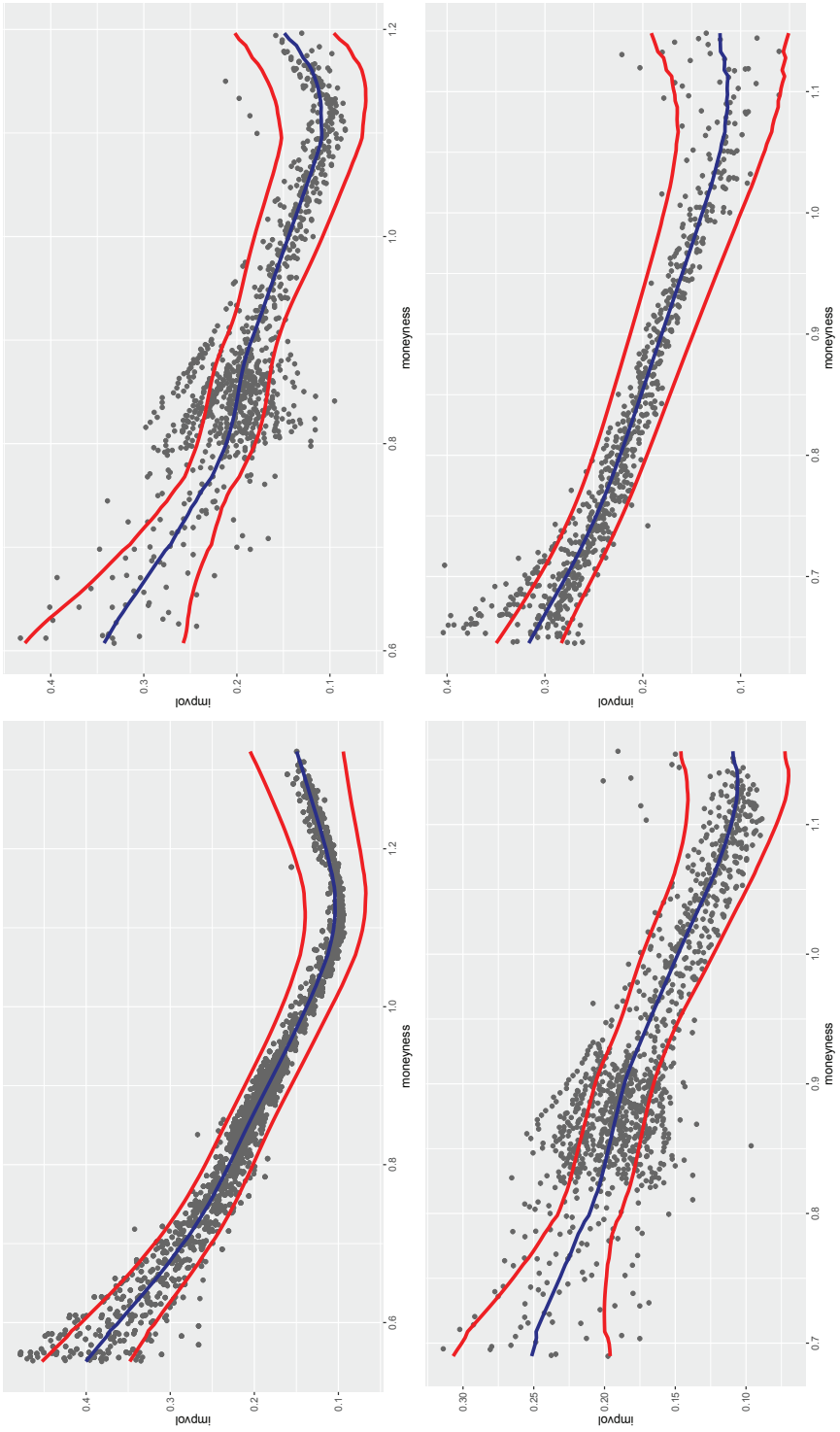

Figure 2.4: Fitted implied volatility and bootstrap uniform confidence bands for 4 (L)ETFs on S&P500; τ: 0.6 years

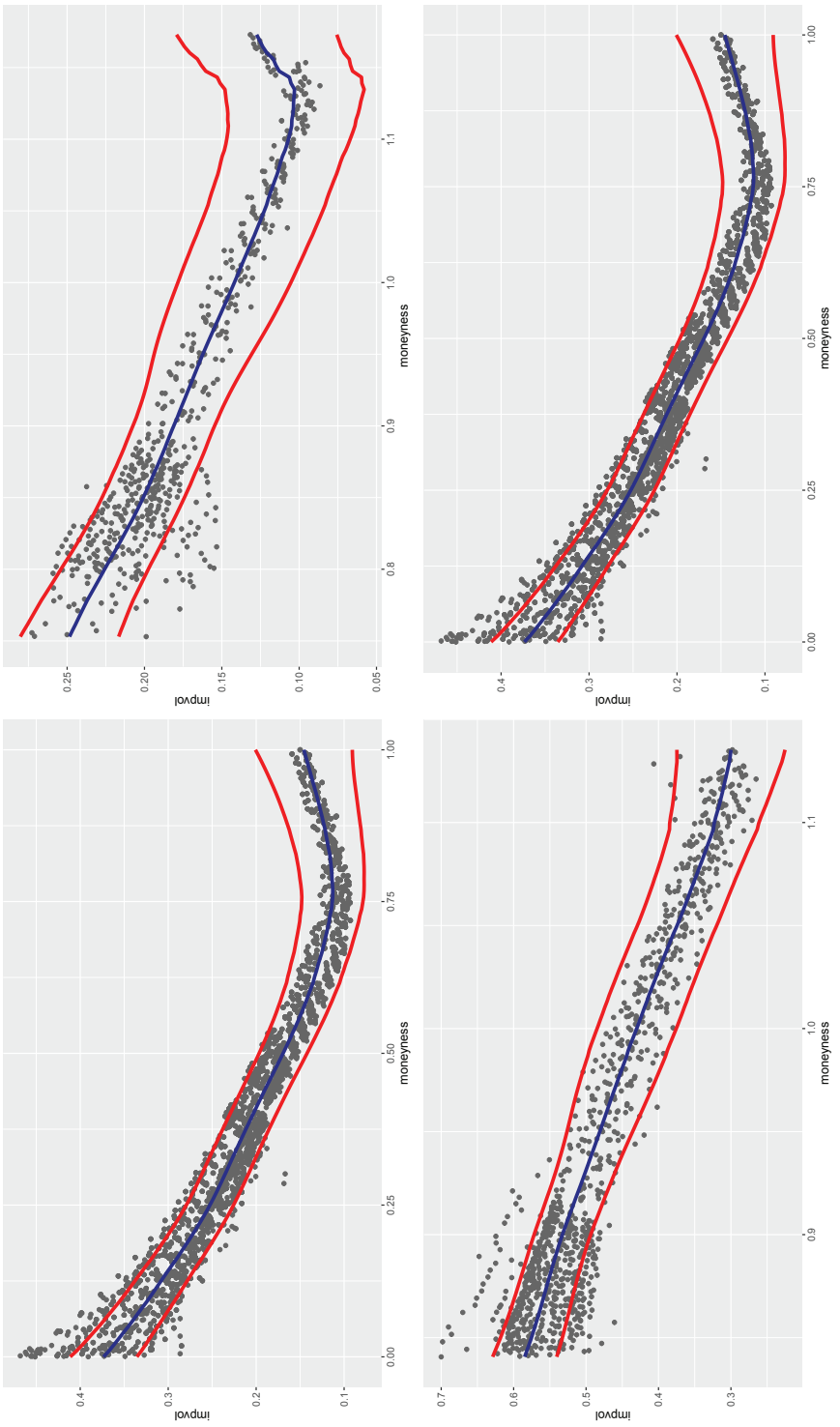

Figure 2.5: Fitted implied volatility and bootstrap uniform confidence bands for 4 (L)ETFs on S&P500; τ: 0.7 years

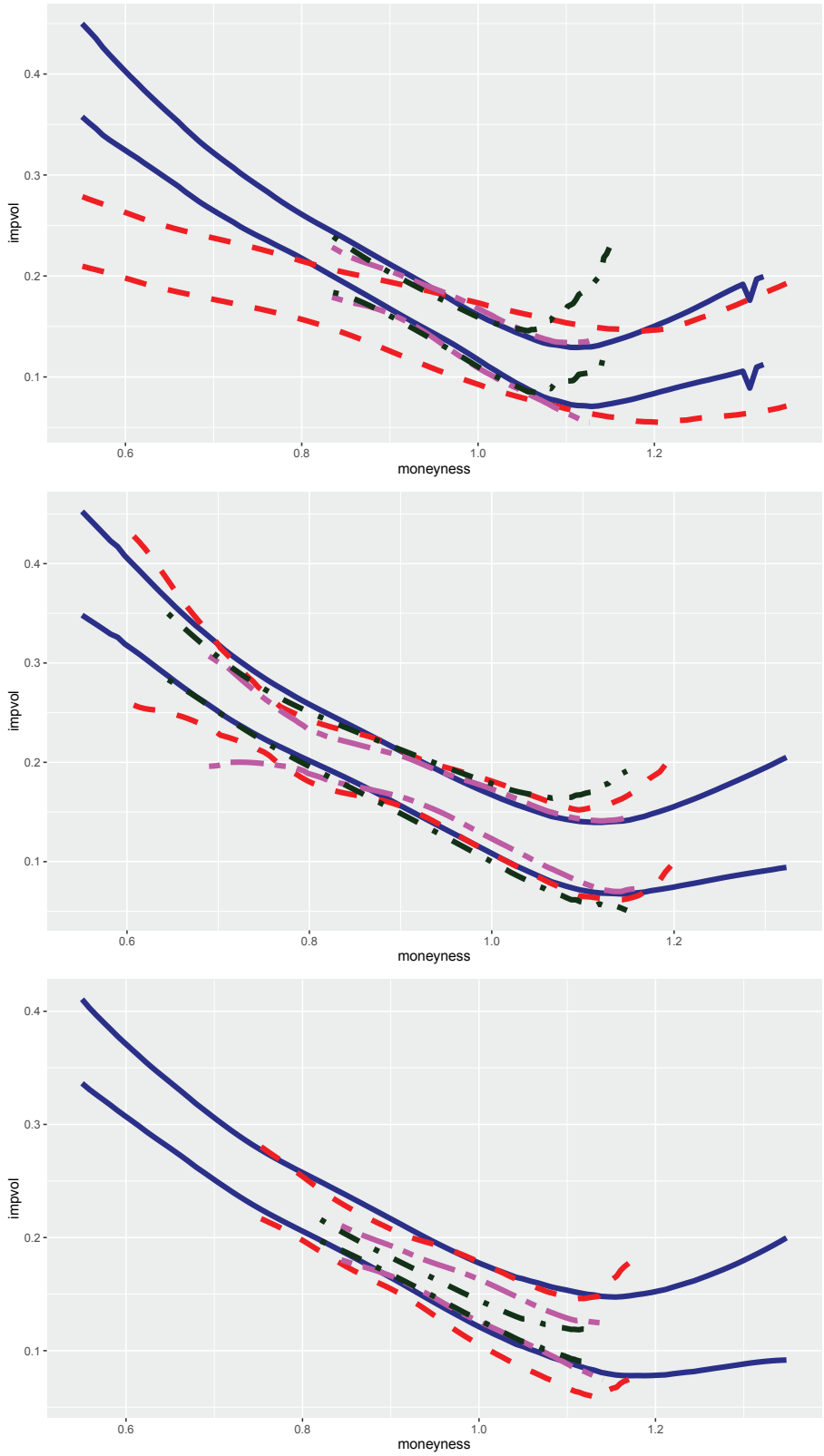

Figure 2.6: Combined uniform bootstrap confidence bands for SPY, SSO, UPRO and SDS after moneyness scaling

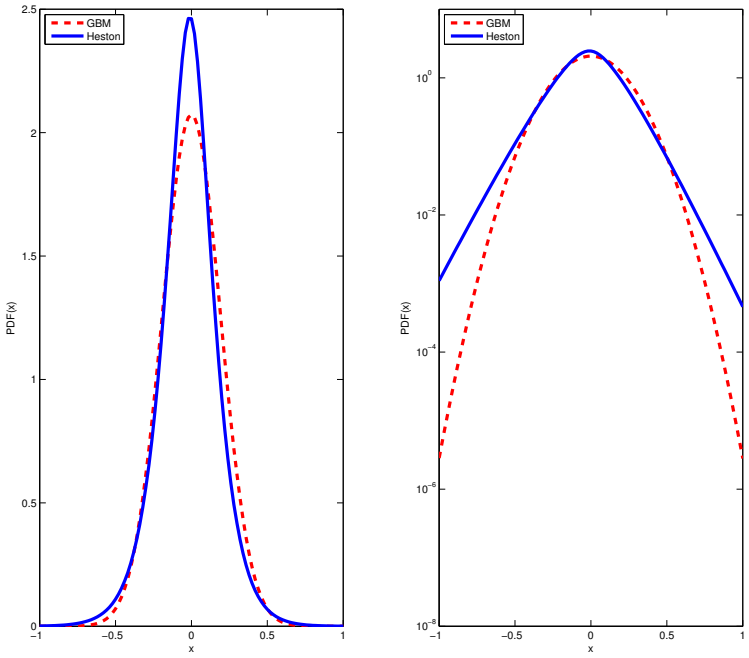

Figure 2.7: probability density functions of log-returns implied by Heston and GBM, 20160201

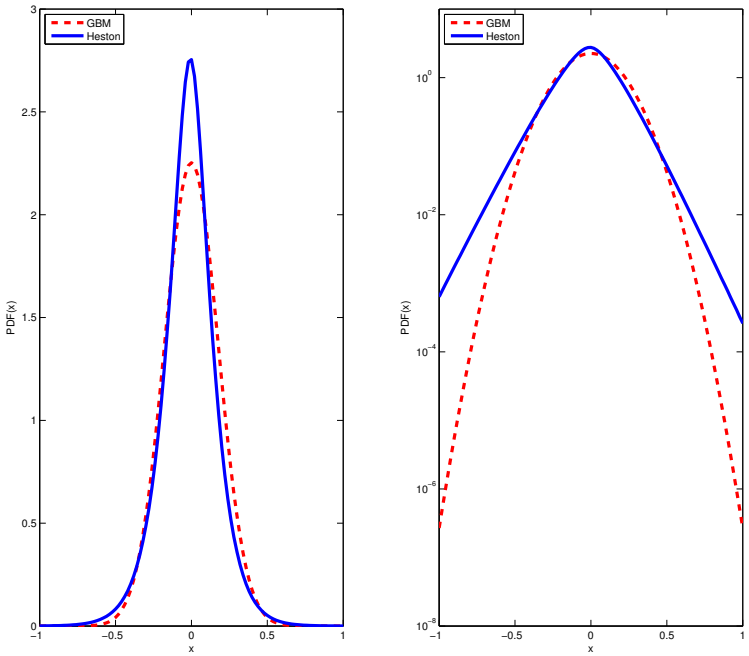

Figure 2.8: probability density functions of log-returns implied by Heston and GBM, 20160205

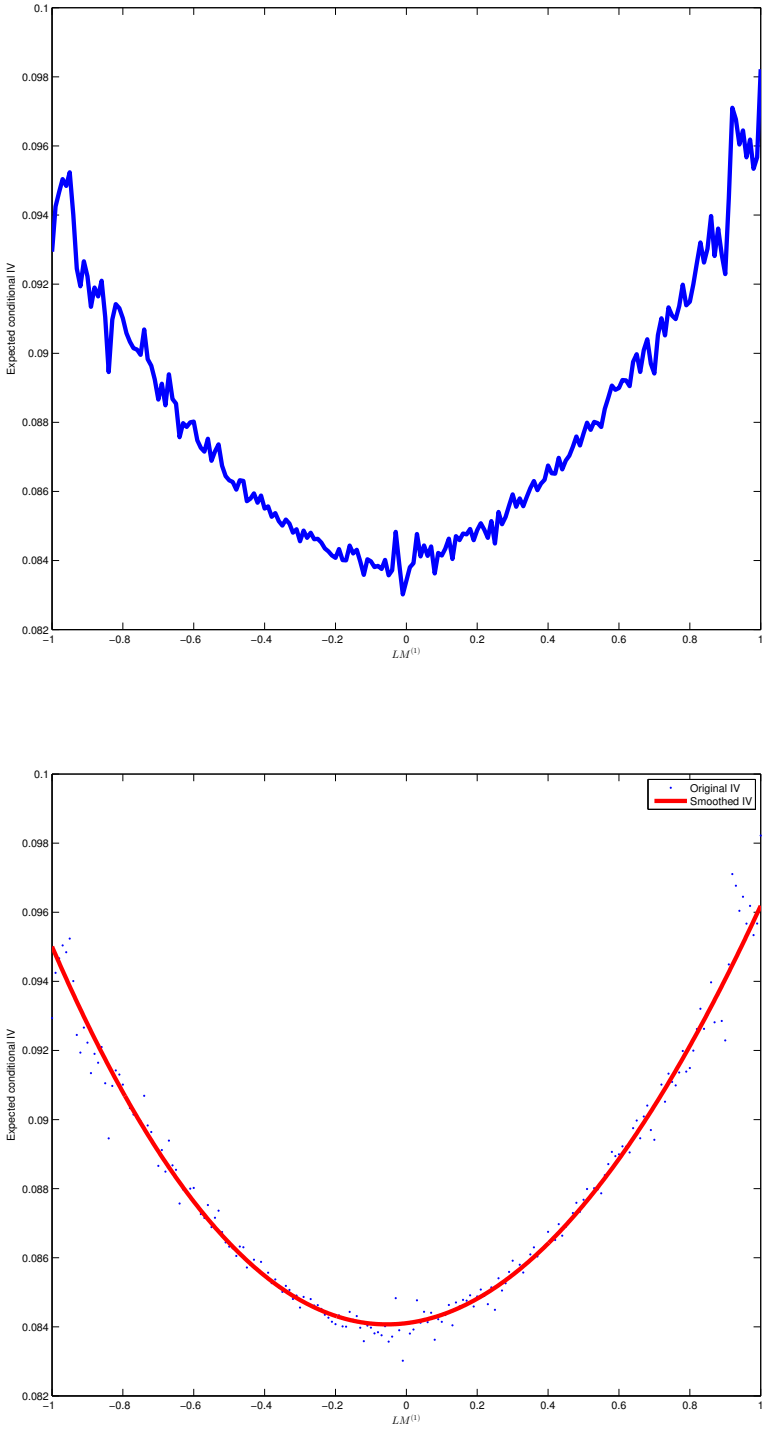

Figure 2.9: Upper panel: estimated value of (2.12); lower panel: smoothed estimate

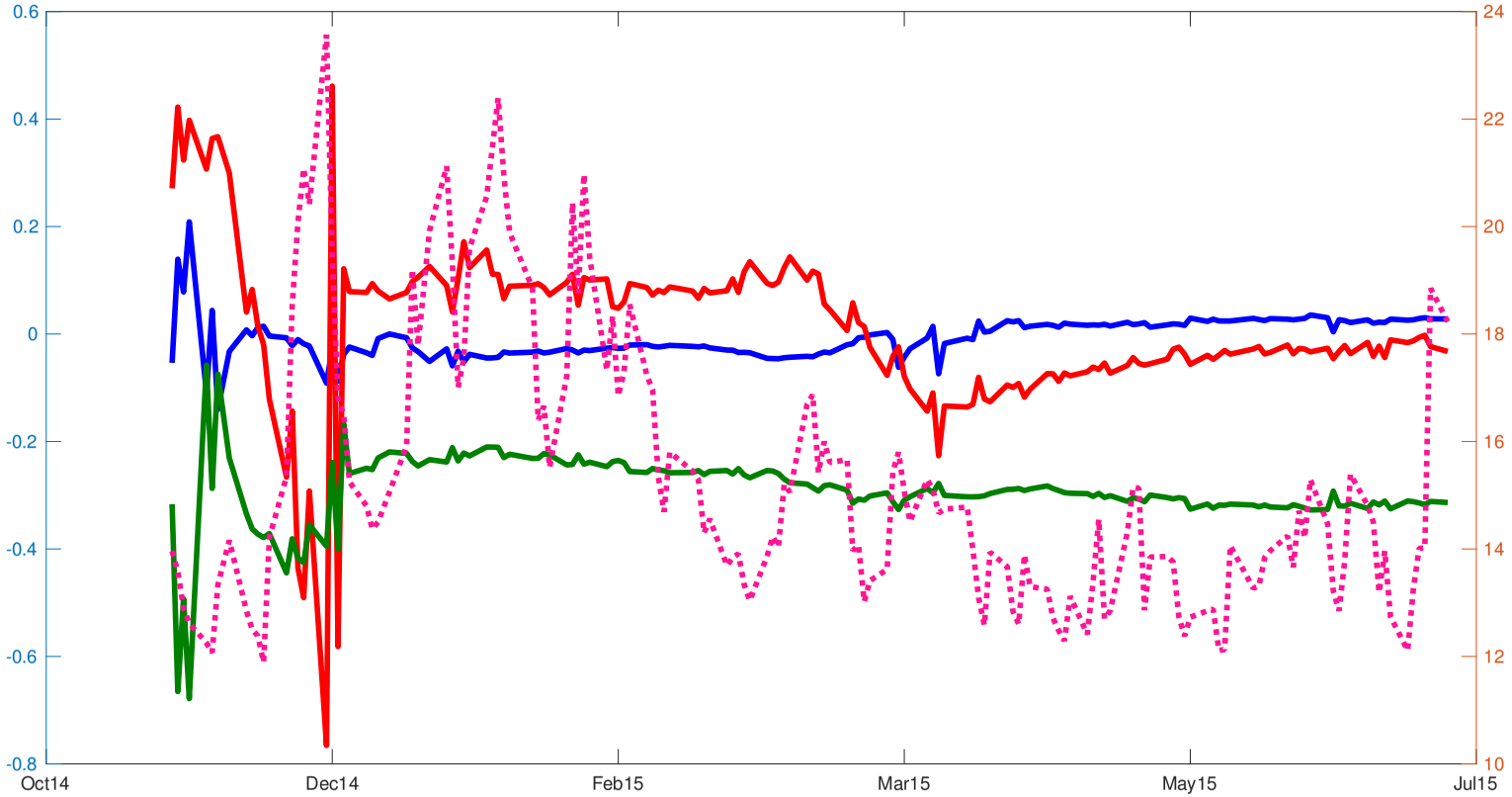

Figure 2.10: Time dynamics of $\widehat{\mathscr{Z}_{t,1}}$, $\widehat{\mathscr{Z}_{t,2}}$, $\widehat{\mathscr{Z}_{t,3}}$, VIX index

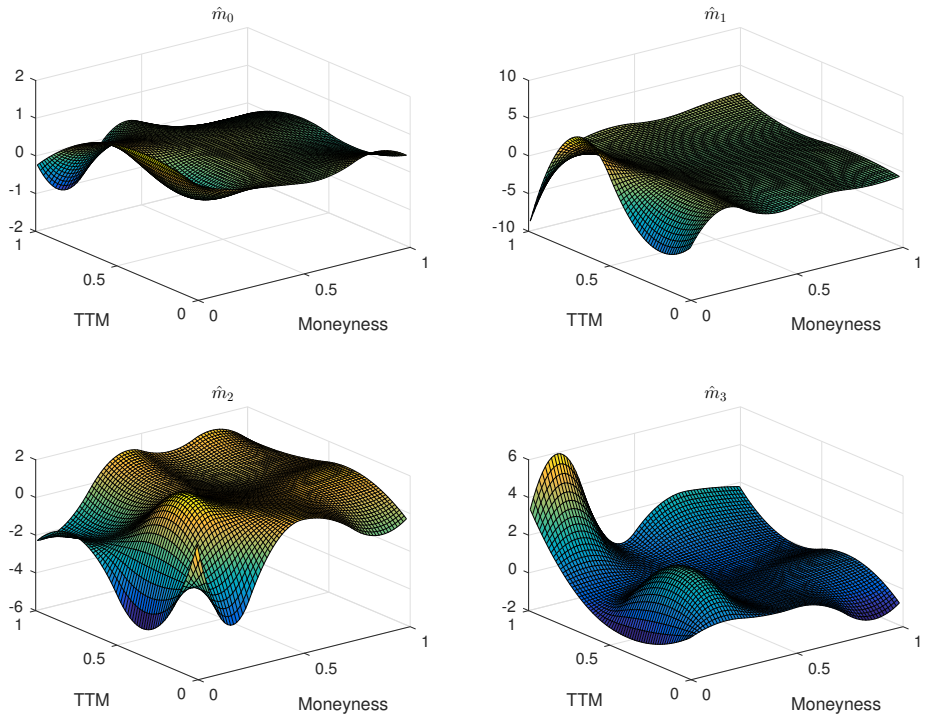

Figure 2.11: Factor functions $\hat{m}_0$, $\hat{m}_1$, $\hat{m}_2$, $\hat{m}_3$ for SPY option

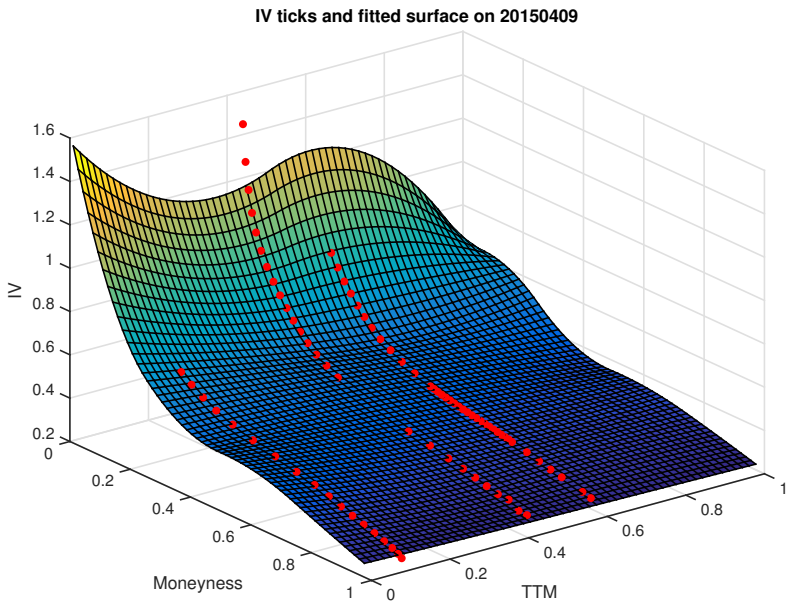

Figure 2.12: Implied volatility real-data "strings" and the DSFM-fitted surface on 20150409

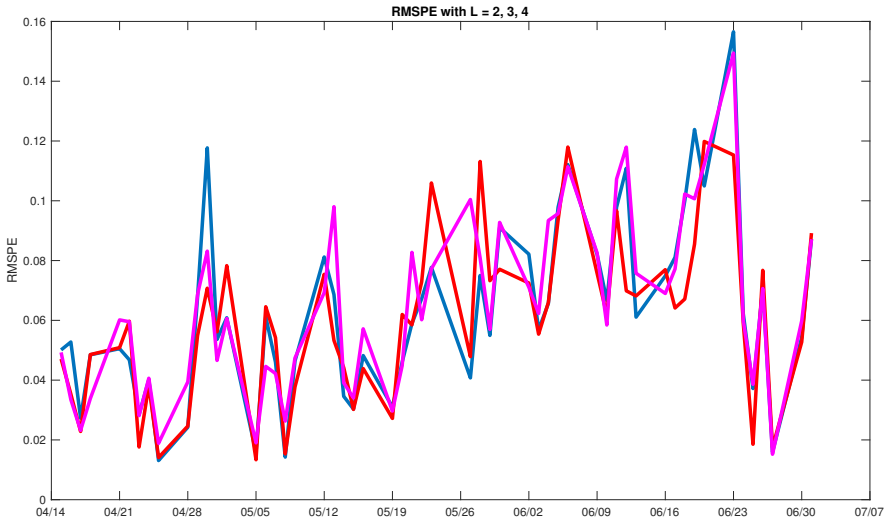

Figure 2.13: RMSPE for $L = 2$, $L = 3$, $L = 4$ for the year 2015

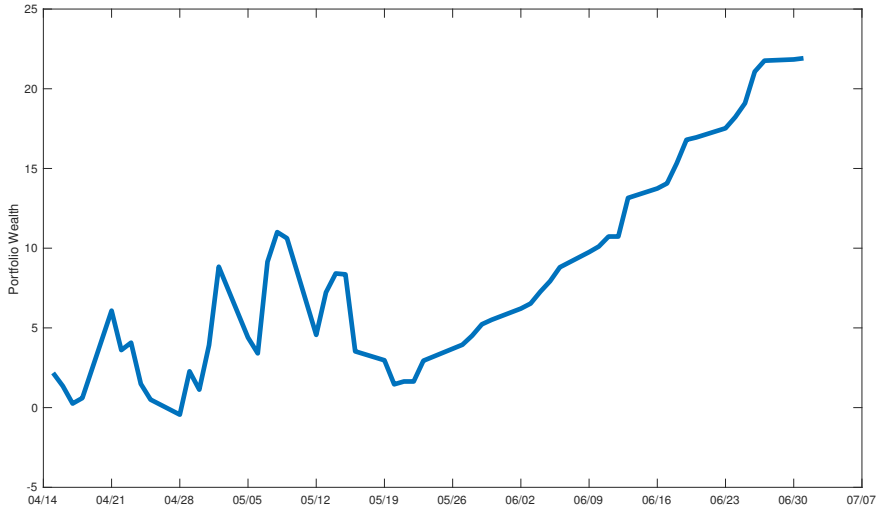

Figure 2.14: Cumulative performance of the dynamic strategy for the year 2015

Chapter 3

Quantifying systemic risk with factor copulae

3.1 Introduction

Systemic risk is a very important aspect of economic risk which played a significant role in the events of the financial crisis. It continues to be an extremely relevant topic today. An important question is how systemic risk can be quantified. The notion of systemic risk and a macroprudential approach, relevant to financial stability and the functioning of financial markets, have gained significant attention of regulators, financial analysts and academic researchers.

The Financial Stability Board (FSB) together with the Basel Committee on Banking Supervision (BCBS) developed a methodology to select global systemically important banks (G-SIBs) and attribute them to categories ("buckets"). Depending on the "buckets", additional common equity loss absorbency is prescribed in terms of a percentage of risk-weighted assets. This methodology is mainly based on the so-called "indicator-based measurement approach" which is based on a number of indicators postulated to capture the level of global systemic importance. Among them are such indicators as bank size measured by total exposures, interconnectedness, substitutability and complexity, see Basel Committee on Banking Supervision (2013). Though these indicators are important, they do not necessarily reflect the global scope of the bank's operations and may suffer from arbitrary weight assignment.

The level of "interconnectedness" risk in a system of financial institutions is studied via the factor copula model which is able to identify extreme tail dependence between the institutions and therefore accurately captures the level of risk during financial crises. Simultaneously network risk measures based on eigenvector cen-

trality analysis of similarity matrices are introduced. Measures based on the tail dependence structure outperform those based on a correlation structure. It is found that the classification obtained from the factor copula model and network risk measures differs from the official classification by the FSB and the BCBS, which reflects their focus on the "too big to fail" approach as opposed to "too connected to fail".

3.2 Network centrality analysis

3.2.1 Eigenvector centrality

Eigenvector centrality analysis allows to identify the most “important“ vertices (nodes) in networks. Using the adjacency matrix of a network (graph), we can track the neighbors for each node v_i. Let $\gamma(v_i)$ denote node centrality and define the centrality of a node proportional to the sum of its neighbors' centralities:

$$\gamma(v_i) = \frac{1}{\lambda} \sum_{j=1}^{N} a_{j,i} \gamma(v_j) \tag{3.1}$$

where $a_{j,i}$ are the elements of the centrality matrix A and λ is a fixed constant. Letting $\Gamma = (\gamma(v_1), \gamma(v_2), ..., \gamma(v_N))^T$ as the centrality vectors for all nodes, we can restate the above equation as

$$\lambda \Gamma = A\Gamma \tag{3.2}$$

Eq. 3.2 indicates that Γ is an eigenvector of the adjacency matrix, A, and λ is the corresponding eigenvalue. In fact, if we choose to impose a positivity constraint on the centralities' vector Γ, this is the largest eigenvalue of the adjacency matrix A, and the corresponding eigenvector is the vector of network centralities. The central nodes can be selected by ranking the elements in the selected eigenvector.

3.2.2 Adjacency matrix

Measures of similarity

Statistically, similarity can be defined, for example, in terms of correlation. Given random observations x_{it} and y_{jt}, $t = 1, \ldots, T$, T is a time horizon, Pearson correlation coefficient ρ_{ij}^P is a measure of linear association between them:

$$\rho_{ij}^P \stackrel{\text{def}}{=} \frac{\sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j)}{\sqrt{\sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2} \sqrt{\sum_{t=1}^{T} (x_{jt} - \bar{x}_j)^2}} \tag{3.3}$$

Various rank correlation measures such as Spearman rank correlation ρ_{ij}^S or Kendall's τ are measures of monotone (not necessarily linear) association between two variables. For example, define R_i, R_j as $T \times 1$ vectors of ranks of x_{it}, x_{it}, then:

$$\rho_{ij}^S \stackrel{\text{def}}{=} \frac{\sum_{t=1}^T \{R_{it} - 0.5(T+1)\}\{R_{jt} - 0.5(T+1)\}}{\sqrt{\sum_{t=1}^T \{R_{it} - 0.5(T+1)\}^2}\sqrt{\sum_{t=1}^T \{R_{jt} - 0.5(T+1)\}^2}}. \tag{3.4}$$

Statistical dependence is determined through joint distributions other than those which can be completely obtained just from marginal distributions. Of particular interest are extreme or tail dependencies, because they allow to measure the level of risk in the financial markets during market crashes more efficiently than association measures. Copula functions are flexible and efficient instruments which allow to set a wide range of dependency between random variables with various marginals.

Given d dimensions, a copula is an d-dimensional joint distribution with $U[0,1]$-uniform marginals. According to the Sklar's theorem, if C is a copula and $F_{X_1}, \ldots, F_{X_d}$ are continuous marginal distributions of $X_1, \ldots, X_d$, then one can uniquely construct a joint distribution $F(x_1, \ldots, x_d) = C\{F_{X_1}(x_1), \ldots, F_{X_d}(x_d)\}$. For details, see McNeil et al. (2015).

Extreme or tail dependence can be explicitly defined given a specific copula. These measures gauge the strength of dependence in the tails of a bivariate distribution. To be precise, the coefficients of lower and upper tail dependence are defined as follows:

$$\begin{aligned}
\Lambda_{ij}^L &\stackrel{\text{def}}{=} \lim_{q \to 0^+} \mathsf{P}(X_j \leq F_j^{-1}(q) | X_i \leq F_i^{-1}(q)), \\
&= \lim_{q \to 0^+} \frac{\mathsf{P}(X_i \leq F_i^{-1}(q), X_j \leq F_j^{-1}(q))}{q}, \qquad (3.5) \\
\Lambda_{ij}^U &\stackrel{\text{def}}{=} \lim_{q \to 1^-} \mathsf{P}(X_j > F_j^{-1}(q) | X_i > F_i^{-1}(q)), \\
&= \lim_{q \to 1^-} \frac{\mathsf{P}(X_i > F_i^{-1}(q), X_j > F_j^{-1}(q))}{1-q}. \qquad (3.6)
\end{aligned}$$

Alternatively, as proposed by Schmidt et al. (2006), tail dependence can be estimated by means of empirical tail copulas. This allows to estimate tail dependence coefficients in a non-parametric setting. The marginal distributions are modeled using empirical distribution functions which could to avoid misspecification due to possible wrong parametric fit of the marginal distributions. The non-parametric

estimators for (3.5), (3.6) are written as follows:

$$\widehat{\Lambda}_{ij}^{L} \approx \frac{1}{k}\sum_{t=1}^{T}\mathbf{I}_{\left\{R_i^{(t)}\leq kx_i, R_j^{(t)}\leq kx_j\right\}}, \tag{3.7}$$

$$\widehat{\Lambda}_{ij}^{U} \approx \frac{1}{k}\sum_{t=1}^{T}\mathbf{I}_{\left\{R_i^{(t)}>T-kx_i, R_j^{(t)}>T-kx_j\right\}}. \tag{3.8}$$

The parameter k (threshold) is chosen via a plateau-finding algorithm which corresponds to balancing bias and variance. The estimators are shown to have asymptotically normal distribution under both known and unknown marginal distributions. The details can be found in Schmidt et al. (2006).

The similarity matrix is constructed from pairwise tail dependence coefficients Λ_{ij}^{L}, Λ_{ij}^{U} or the empirical counterparts $\widehat{\Lambda}_{ij}^{L}$, $\widehat{\Lambda}_{ij}^{U}$ for lower and upper tail dependence, respectively. Once similarity between financial variables has been determined, further analysis is necessary to determine the network (adjacency) structure of the underlying system.

Breakpoint analysis

Relevant estimation of the adjacency matrix is required for eigenvector centrality analysis. One way to achieve this is to convert a similarity matrix into a binary adjacency structure. The method of Ng (2006) proposes a breakpoint analysis framework which is based on uniform "spacings" analysis. The problem of testing cross-section correlation is turned into a problem of testing uniformity and nonstationarity. A subset of nonzero correlations can be determined by minimizing a sum of square residuals.

The idea of uniform spacings can be generalized to any similarity matrix measure as long as its elements can be assumed $U[0,1]$-distributed. To be precise, given a $N \times N$ similarity matrix p_{ij}, $i, j = 1, \ldots, N$, breakpoint determination is achieved via several steps:

1. sort cross-sectional similarities into an ordered vector $\mathbf{p} = (p_1, p_2, ..., p_n)$, where p_1 is the smallest one and p_n is the largest one, $n = N(N-1)/2$,

2. perform a uniform transformation of $\mathbf{p}$ via the standard normal cumulative distribution function:

$$\mathbf{\Phi} = \left(\Phi\left(\sqrt{T}|p_1|\right), \Phi\left(\sqrt{T}|p_2|\right), \ ... \ , \Phi\left(\sqrt{T}|p_n|\right)\right), \tag{3.9}$$

3. calculate the "spacings" $\Delta\phi_j = \Phi\left(\sqrt{T}|\rho_j|\right) - \Phi\left(\sqrt{T}|\rho_{j-1}|\right)$,

4. perform the optimization

$$\hat{\theta} = \arg\min_{\theta \in \mathbb{R}} f_n(\theta), \tag{3.10}$$

where $f_n(\theta) = \sum_{j=1}^{\lceil \theta n \rceil} \left(\Delta\phi_j - \mu_s\right)^2 + \sum_{j=\lceil \theta n \rceil + 1}^{n} \left(\Delta\phi_j - \mu_L\right)^2$ and $\mu_s = \frac{1}{\lceil \theta n \rceil} \sum_{j=1}^{\lceil \theta n \rceil} \Delta\phi_j$, $\mu_L = \frac{1}{\lceil n(1-\theta) \rceil} \sum_{j=\lceil \theta n \rceil + 1}^{n} \Delta\phi_j$, $\lceil \theta n \rceil$ is the integer part of θn.

As pointed out by Chan-Lau et al. (2016), nodes' characteristics may affect the nature of their connections and the network characteristics. Therefore a weighted adjacency matrix can be constructed. Accounting for skewness and kurtosis in centrality analysis is useful for finding the "conditioning" factors which cause a higher risk level for the entire financial network. This is also confirmed by the findings of Girardi and Tolga Ergün (2013) in their CoVaR study of systemic risk. Denote by Q a diagonal matrix with diagonal entries q_{jj} being a specific characteristic weight for the jth financial institution, then the adjusted adjacency matrix A^* is defined by

$$A^* \stackrel{\text{def}}{=} QAQ. \tag{3.11}$$

As a result, $\left\lfloor \widehat{\theta} n \right\rfloor$ yields an optimal break location achieving a minimum total sum of variances from two subgroups to make the correlations in the given group as homogeneous as possible.

3.3 Factor copulas

3.3.1 General theory

Copulas in general are flexible tools of modelling multivariate distributions which allow for separate modelling of marginal distributions and the dependence structure. The Sklar's theorem postulates that every multivariate distribution can be represented via the corresponding marginal distributions and a copula.

This property allows to construct a wide range of dependence types for random variables which are converted to $U(0,1)$-uniform ones. This is done to guarantee that a copula has uniform univariate marginal distributions.

Factor copula models go a step further from other copula types to address the issue of high dimensionality and polynomial-time complexity in copula parameter estimation. Given d marginal distributions, usual copula constructions (e.g., direct multivariate copulas, vines) involve estimating $\mathcal{O}(d^2)$ parameters. Factor copulas

allow to do parameter estimation in linear time: e.g., compared to vine pair-copula models, they reduce the number of parameters to be estimated to $\mathcal{O}(d)$, see Krupskii and Joe (2013).

A general multivariate factor copula model assumes a linear dependence structure of d observed variables Z and p "latent" variables W:

$$Z_j = \theta_{j|1} W_1 + \ldots + \theta_{j|p} W_p + \psi_j \varepsilon_j, \quad j = 1, \ldots, d. \tag{3.12}$$

This representation extends the linear Gaussian factor model $R = AA^\top + \Psi^2$ where A is a $d \times p$ matrix of loadings with $1 \leq p < d$ and Ψ^2 is a diagonal matrix of "specific" variances $\psi_1^2, \ldots, \psi_d^2$.

In a special univariate case, the representation (3.12) assumes the form:

$$Z_j = \theta_{j|1} W + \sqrt{1 - \theta_{j|1}^2}\, \varepsilon_j, \quad j = 1, \ldots, d. \tag{3.13}$$

In the factor copula model, the copula-dependent uniform random variables $u_j \stackrel{\text{def}}{=} F_{Z_j}(z_j)$, $j = 1, \ldots, d$, obtained from the marginal transformation of Z_j in $Z \stackrel{\text{def}}{=} (Z_1, \ldots, Z_d)^T$ are assumed to be conditionally independent given p latent variables V. The factor copula expression is then derived via the *mixture families* approach. Assume $p = 1$ (one-factor case), define $U \stackrel{\text{def}}{=} (U_1, \ldots, U_d)^T$, V, all $U(0,1)$, i.i.d., then:

$$\begin{aligned} C_v(u_1, \ldots, u_d) &= F(Z_1, \ldots, Z_d) \\ &= \int_D F_{Z|V}(z|v) \mathrm{d}F_V(v) \\ &= \int_D \prod_{j=1}^{d} F_{Z_j|V}(z_j|v) \mathrm{d}F_V(v) \\ &= \int_D \prod_{j=1}^{d} C_{F_Z(Z_j)|V}(F_Z(z_j)|v) \mathrm{d}v \\ &= \int_D \prod_{j=1}^{d} C_{U_j|V}(u_j|v) \mathrm{d}v, \end{aligned} \tag{3.14}$$

denotes a multivariate factor copula with conditionally independent marginals $U_1, \ldots, U_d$, given the latent factor V; here $D \stackrel{\text{def}}{=} [0,1]$, the first and fourth equality come from the Sklar's theorem and uniformity, the third from the independence assumptions.

The expression (3.13) allows to generate different types of copulas given the distributions of W and ε. Oh and Patton (2015) use normal, t and Skew t distributions to

make the copula exhibit asymmetric and tail dependence. These copulas normally do not have closed form. In a simple example with W and ε both being $\mathrm{N}(0,1)$, the resulting copula is Gaussian. It follows that (see Appendix 4.7)

$$C_{U_j|V}(u_j|v) = \Phi\left(\frac{\Phi^{-1}(u_j) - \theta_{j|v}\Phi^{-1}(v)}{\sqrt{1-\theta_{j|v}^2}}\right). \tag{3.15}$$

The resulting expression for (3.14) is then

$$C_w(u_1,\ldots,u_d) = \int_D \prod_{j=1}^{d} \left\{(\Phi^{-1}(u_j) - \theta_{j|w}w)/\sqrt{1-\theta_{j|w}^2}\right\} \varphi(w)\mathrm{d}w. \tag{3.16}$$

In general, the conditional pair-copula for the factor model, given independent uniformly distributed random variables $V = v$, U_j, takes the form (see Appendix 4.8)

$$C_{U_j|V}(u_j|v) = F_{\varepsilon_j}\left(\frac{F_{Z_j}^{-1}(u_j) - \theta_{j|1}v}{\sqrt{1-\theta_{j|1}^2}}\right), \tag{3.17}$$

see also McNeil et al. (2015), Kalemanova et al. (2007). Here W, ε_j can have arbitrary continuous distributions, the distribution F_{Z_j} is obtained from the convolution of $\theta_{j|1}W$ and $\sqrt{1-\theta_{j|1}^2}\varepsilon_j$, according to the form (3.13).

3.3.2 Factor copula under particular distributions

Specific variations of (3.17) are obtained by using parametric continuous distributions for the common factor W and the idiosyncratic shock ε_j, respectively. Some common examples, see McNeil et al. (2015), include the so-called double-t and the double-GH copulas where both W and ε_j follow univariate t_ν and generalized hyperbolic (GH) distributions, respectively. If one utilizes the representation (3.14)-(3.17) for parameter estimation, one has to numerically compute F_{Z_j} (via convolution) and its inverse at a particular point in every iteration. This makes the computation prohibitively slow and the model impossible to use for practical purposes.

One can address this problem by using distributions which possess stability under convolution as well as fit financial data well. Among such distributions the family of stable distributions is found which, for specific values of their parameters, asymptotically exhibit power law behaviour in the tails ("heavy-tailed" distributions).

Also in this group we find the class of GH distributions which are closed under convolution given certain constraints on their parameters. As was shown in previous

research by Borak et al. (2010), statistical tests such as Kolmogorov and Anderson-Darling goodness-of-fit statistics show that two subclasses of the *GH* distribution, the hyperbolic and the normal inverse Gaussian (*NIG*) distributions provide the best model for financial data. The double-*NIG* copula approach was applied by Kalemanova et al. (2007) for synthetic CDO pricing. Explicit dependence of the parameters of the convolution distribution F_{Z_j} on the factor loading parameter $\theta_{j|1}$ had to be introduced in order to perform the convolution.

A drawback of *GH* distributions is that they allow for nonzero tail dependence in the factor copula framework only under restrictive assumptions. An alternative approach by Oh and Patton (2016) derives the factor copula likelihood via a change of variables so that no convolution is necessary. The likelihood function is computed using numerical integration and bilinear interpolation methods.

3.3.3 Tail dependence for factor copulas

For a factor copula generated by a linear structure (3.13) tail dependence coefficients (3.5) and (3.6) can be derived in explicit form.

Proposition 3.3.3.1. Let the factor copula be generated by the linear factor structure (3.13). Also let F_W and F_{ε_j} have regularly varying tails with a common tail index $\alpha > 0$ so that $\mathsf{P}(W < -s) = \mathsf{P}(W > s) = A_W s^{-\alpha}$, $\mathsf{P}(\varepsilon_j < -s) = \mathsf{P}(\varepsilon_j > s) = A_\varepsilon s^{-\alpha}$ as $s \to \infty$, $A_W > 0$, $A_\varepsilon > 0$.

Then it follows that

$$\Lambda^L_{ij} = \frac{A_W \theta^\alpha_{i|1}}{A_W \theta^\alpha_{i|1} + A_\varepsilon (1-\theta^2_{i|1})^{\alpha/2}} \tag{3.18}$$

if the following conditions hold: $A_W \theta^\alpha_{i|1}\theta^\alpha_{j|1} + A_\varepsilon(1-\theta^2_{i|1})^{\alpha/2}\theta^\alpha_{j|1} > A_W\theta^\alpha_{j|1}\theta^\alpha_{i|1} + A_\varepsilon(1-\theta^2_{j|1})^{\alpha/2}\theta^\alpha_{i|1}$ and simultaneously $\theta_{i|1} < \theta_{j|1}$ or $A_W\theta^\alpha_{i|1}\theta^\alpha_{j|1} + A_\varepsilon(1-\theta^2_{i|1})^{\alpha/2}\theta^\alpha_{j|1} < A_W\theta^\alpha_{j|1}\theta^\alpha_{i|1} + A_\varepsilon(1-\theta^2_{j|1})^{\alpha/2}\theta^\alpha_{i|1}$ and simultaneously $\theta_{i|1} > \theta_{j|1}$. On the other hand, it holds that

$$\Lambda^L_{ij} = \frac{A_W \theta^\alpha_{j|1}}{A_W \theta^\alpha_{j|1} + A_\varepsilon (1-\theta^2_{j|1})^{\alpha/2}} \tag{3.19}$$

if the following conditions hold: $A_W \theta^\alpha_{i|1}\theta^\alpha_{j|1} + A_\varepsilon(1-\theta^2_{i|1})^{\alpha/2}\theta^\alpha_{j|1} < A_W\theta^\alpha_{j|1}\theta^\alpha_{i|1} + A_\varepsilon(1-\theta^2_{j|1})^{\alpha/2}\theta^\alpha_{i|1}$ and simultaneously $\theta_{i|1} < \theta_{j|1}$ or $A_W\theta^\alpha_{i|1}\theta^\alpha_{j|1} + A_\varepsilon(1-\theta^2_{i|1})^{\alpha/2}\theta^\alpha_{j|1} > A_W\theta^\alpha_{j|1}\theta^\alpha_{i|1} + A_\varepsilon(1-\theta^2_{j|1})^{\alpha/2}\theta^\alpha_{i|1}$ and simultaneously $\theta_{i|1} > \theta_{j|1}$.

Proof. See Appendix 4.9. □

Proposition 3.3.3.2. Let the factor copula be generated by the linear factor structure (3.13). Also let F_W and F_{ε_j} be $t(\mu, \sigma, \nu)$ and $t(\nu)$, then

$$A_W = \frac{(\nu\sigma^2)^{\frac{\nu+1}{2}}}{\nu^{3/2}\sigma B(n/2, 1/2)}, \tag{3.20}$$

where $B(\cdot,\cdot)$ is the beta function.

Proof. See Appendix 4.10. □

3.3.4 Copula parameter estimation

Estimation of copula parameters with likelihood methods often involves quantities which do not have closed form, therefore one has to use approximative numerical methods. The likelihood function for the factor copula can be derived via direct differentiation of the integrand in 3.14. Alternatively, one can proceed by differentiating an absolutely continuous joint distribution function $F_{\mathbf{Z}}$ with strictly increasing, continuous marginal distribution functions $F_{Z_1}, \ldots, F_{Z_d}$, which generates an implicit copula $C(u_1, \ldots, u_d)$ with the corresponding density, see McNeil et al. (2015),

$$c(u_1, \ldots, u_d) = \frac{f_{\mathbf{Z}}(F_{Z_1}^{-1}(u_1), \ldots, F_{Z_d}^{-1}(u_d))}{f_{Z_1}(F_{Z_1}^{-1}(u_1))\cdot, \ldots, \cdot f_{Z_d}(F_{Z_d}^{-1}(u_d))}, \tag{3.21}$$

where $f_{\mathbf{Z}}$ is the joint density of $Z_1, \ldots, Z_d$; F_{Z_j}, f_{Z_j}, $j = 1, \ldots, d$ are the marginal cumulative distribution function and density of Z_j, respectively. Following Oh and Patton (2016), it can be shown that $f_{\mathbf{Z}}$, F_{Z_j} and f_{Z_j} take the following forms, assuming 3.23:

$$f_{\mathbf{Z}}(x_1, \ldots, x_d) = \int_0^1 \prod_{j=1}^{d} f_{\varepsilon_j}\left(\frac{x_j - \theta_{j|i} F_{Z_i}^{-1}(v)}{\sqrt{1-\theta_{j|i}^2}}\right) dv,$$

$$F_{Z_j}(x_j) = \int_0^1 F_{\varepsilon_j}\left(\frac{x_j - \theta_{j|i} F_{Z_i}^{-1}(v)}{\sqrt{1-\theta_{j|i}^2}}\right) dv,$$

$$f_{Z_j}(x_j) = \int_0^1 f_{\varepsilon_j}\left(\frac{x_j - \theta_{j|i} F_{Z_i}^{-1}(v)}{\sqrt{1-\theta_{j|i}^2}}\right) dv.$$

One-dimensional numerical integration is performed to determine the integral on the interval $[0, 1]$ in (3.21). Krupskii and Joe (2013) implement the Gauss-Legendre

quadrature for 1- and 2-factor copula models. A quadrature rule approximates the following definite integral on a suitable domain D:

$$\int_D f(x)dx \approx \sum_{k=1}^{q} \omega_k f(x_k), \tag{3.22}$$

where q is the number of quadrature points, x_k are the quadrature points or nodes and ω_k are the quadrature weights. The expressions for ω_k for different quadrature rules can be found, e.g., in Abramowitz and Stegun (1965). According to Joe (2015), the number of quadrature points q around 20-30 per dimension is usually adequate for the maximum likelihood estimate to be numerically stable. In the empirical study below $q = 21$ is used.

Numerical computation also has to be used to determine $F_{Z_j}^{-1}$. It would be computationally expensive to determine this quantity in each iteration of the likelihood optimization, therefore an approximative numerical method is used. First, two grids in the intervals $[0,1]$ and $[-1,1]$ for u and $\theta_{j|i}$, respectively, are created. Then, given a pair of values $(u, \theta_{j|i})$, the value of $F_{Z_j}^{-1}(u;\theta_{j|i})$ can be determined via a root-searching algorithm for the problem $F_{Z_j}(x;\theta_{j|i}) - u = 0$ by solving for x. Given a 2-dimensional rectilinear grid of $F_{Z_j}^{-1}(u;\theta_{j|i})$ values, one can perform bilinear interpolation to determine the values of $F_{Z_j}^{-1}$ in each MLE iteration. The matrix $F_{Z_j}^{-1}(u;\theta_{j|i})$ is computed only once prior to estimation, which significantly saves computational effort.

3.3.5 Copula simulation and portfolio Value-at-Risk (PVaR)

Generation of copula-dependent random numbers given the estimated factor copula parameters $\widehat{\theta}_{j|i}$ is an essential step for PVaR calculation. A straightforward algorithm can be applied to simulate from a 1-factor copula model. Given the number of simulated samples n_{sim} and a forecast horizon H for the PVaR, a $n_{sim} \cdot H \times N$ array $\mathscr{U}$ is pre-allocated and proceed as outlined in Algorithm 1:

Algorithm 1 1-factor copula simulation

1: **for** $i \leftarrow 1, n_{sim}$ **do**
2: Simulate $v, p_1, \ldots, p_N$ as independent $U(0,1)$-distributed random numbers.
3: Compute $u_j = C_{U_j|V}^{-1}(p_j|v;\widehat{\theta}_{j|i})$, $j = 1,\ldots,N$.
4: Return $(u_1,\ldots,u_N)$.
5: Store $(u_1,\ldots,u_N)$ in the ith row of $\mathscr{U}$.
6: **end for**

The resulting row vectors $(u_1,\ldots,u_N)$ in $\mathscr{U}$ will be a sample from the distribution $C_{z_i}(u_1,\ldots,u_N;\widehat{\theta}_{j|i})$. Copula-dependent random numbers in the second step of Algo-

rithm 1 are determined via numeric inversion of (3.24) as mentioned in the previous section.

Given $\mathscr{U}$, in the last step of Algorithm 2 the autocorrelation and heteroscedasticity observed in the original asset returns are re-introduced back into the copula-dependent uniform random values for PVaR calculation.

The "central nodes" obtained from Section 3.2.1 are the financial institutions with a high level of "connectedness" to the rest of SIFIs. So they can be perceived "close" to the latent factors W in the factor copula model in the distributional sense. That is, if we control for the network effect of these institutions, we achieve approximate conditional "independence" in the network.

Then the factor representation (3.13) assumes the following form:

$$Z_j = \theta_{j|i} Z_i + \sqrt{1-\theta_{j|i}^2}\varepsilon_j, \;\; j = 1,\ldots,i-1,i+1,\ldots,N, \tag{3.23}$$

where Z_i are "non-central" SIFIs, N is the total number of SIFIs, i is the central node index.

The corresponding expression for the copula (3.17) is then:

$$C_{U_j|Z_i}(u_j|z_i) = F_{\varepsilon_j}\left(\frac{F_{Z_j}^{-1}(u_j) - \theta_{j|i} z_i}{\sqrt{1-\theta_{j|i}^2}}\right), \tag{3.24}$$

The level of systemic risk in the group of the "non-central" SIFIs Z_j given a particular "central" SIFI Z_i is then quantified by the factor-copula-based portfolio Value-at-Risk (PVaR) estimated according to the following algorithm:

Algorithm 2 Factor copula PVaR calculation

1: Perform univariate ARMA-GARCH filtering of observed variables Z_j.
2: Derive uniform marginals u_j for each Z_j via marginal cumulative distribution function transformation as stated in 3.3.1.
3: Estimate copula parameters $\theta_{j|i}$ by maximum likelihood.
4: Generate copula-dependent random numbers given the estimates $\widehat{\theta}_{j|i}$.
5: Perform ARMA-GARCH simulation of dependent residuals and calculate the PVaR as 5% or 1%-quantile of the simulated portfolio returns

3.4 Network risk measures

Three network risk measures are introduced based on the adjacency matrix and the factor copula structure model described in earlier sections. These measures are constructed to capture the level of overall systemic risk in a network of financial institutions as follows:

1. Eigenvector centrality based on $p_{ij} = \rho_{ij}^P$
2. Eigenvector centrality based on $p_{ij} = \widehat{\Lambda}_{ij}^L$
3. Singular value norm of $p_{ij} = \Lambda_{ij}^L$ implied by the factor copula model (see sections below).

Note that for a $m \times n$ matrix A the (largest) singular value norm of A is defined as:

$$\begin{aligned} \|A\| &\stackrel{\text{def}}{=} \max_x \|Ax\|_2 \\ \text{s.t.} \quad \|x\|_2 &= 1 \end{aligned}$$

The solution can be derived as:

$$\|A\| = \sqrt{\lambda_{\max}}, \tag{3.25}$$

where $\lambda_{\max}$ is the largest eigenvalue of the positive semidefinite matrix $A^\top A$. For an adjacency matrix A, the singular value norm determines the "magnitude" of the matrix and the level of network risk caused by the level of "connectedness" in the financial system which is generated, e.g., by extreme tail dependence or monotone statistical association.

The third network risk measure is defined as a matrix norm because in the factor copula framework, one of the institutions is inevitably lost during risk calculation given the factor structure (3.13) and we have only $N-1$ institutions in the network instead of N, so we should not directly use eigenvector centrality for risk calculation in this case.

3.5 Empirical analysis

Empirical portfolio VaR analysis is performed based on the factor copula framework. The data for analysis are daily log-returns for 28 systemically important financial institutions (SIFIs) in the period 20070101-20141231. The SIFIs chosen for analysis are the institutions which are regularly selected as systemically by the

Financial Stability Board and the Basel Committee on Banking Supervision in the course of the studied period. Table 3.1 summarizes selected information about the SIFIs.

Figures 3.1, 3.2, 3.3, 3.4 show, respectively, empirical tail dependence matrices $\widehat{\Lambda}_{ij}^{L}$ obtained from the data, adjacency matrices generated according to the breakpoint detection algorithm described in Section 3.2.2, network structures based on these adjacency matrices. The network plots are constructed by means of a force-directed graph drawing algorithm by Fruchterman and Reingold (1991). The data are filtered with a mean-GARCH model before network analysis. Student-t distribution is selected for both the market factor Z_i and the idiosyncratic element ε_j according to goodness-of-fit analysis.

Portfolio Value-at-Risk calculation is performed for each year 2007-2014 under consecutive assumptions of each SIFI being the market factor in the representation (3.13). Then parameter estimation and PVaR calculation are performed as outlined in 3.3.4 and 3.3.5. Once the copula-based PVaR has been determined, network risk analysis with each of the 3 network risk measures outlined in 3.4 is conducted. The goal is to understand which network risk measure most accurately detects the institutions effecting the largest amount of network tail risk.

The results of this analysis are provided in Tables 3.2, 3.3, 3.4. The findings confirm the official classification by the Financial Stability Board and the Basel Committee on Banking Supervision only partially. For instance, in 2007-2008 such SIFIs as Société Générale, Santander, Royal Bank of Scotland are chosen as central nodes but are normally assigned to the least risky "bucket" in the official methodology. Out of the three network-based risk measures assumed, the one based on the singular-value matrix norm of the copula-implied tail dependence matrix is most accurate in selecting the institutions which trigger the largest amount of risk in the financial system as determined by PVaR analysis.

3.6 Conclusions

In this chapter analysis of the network tail risk structure is performed for 28 systemically important financial institutions (SIFIs) selected by the Financial Stability Board and the Basel Committee of Banking Supervision. A factor copula approach is applied which allows to separate out market factors and determine the amount of systemic risk generated by a network where the dynamics of each member is generated by the given factor.

This analysis is relevant for constructing a classification of financial institutions according to their "connectedness". The findings show that the official rankings by the FSB and the BCBS do not always reflect the level of network tail risk in the

system. Furthermore, network risk measures based on network centrality analysis are constructed, all of which confirm these findings.

Factor copula linear structure allows to construct network tail risk matrices explicitly. A singular norm tail matrix risk measure is defined which accurately gauges the level of network risk in a system of financial institutions determined by the factor copula Value-at-Risk measure. This indicator can be generalized via a dynamic structure of factor loadings in the factor copula framework.

Table 3.1: Summary information on systemically important financial institutions (SIFIs)

Index	SIFI	Firm Size	Debt Ratio	Bucket	Country
1	JP MORGAN CHASE	21.506	0.261	4	U.S.
2	BANK OF AMERICA	21.446	0.302	2	U.S.
3	BANK OF NEW YORK MELLON	19.499	0.095	1	U.S.
4	CITIGROUP	21.359	0.300	3	U.S.
5	GOLDMAN SACHS	20.624	0.509	2	U.S.
6	MORGAN STANLEY	20.501	0.417	2	U.S.
7	STATE STREET	19.106	0.153	1	U.S.
8	WELLS FARGO	20.980	0.183	1	U.S.
9	ROYAL BANK OF SCTL	21.588	0.252	1	U.K.
10	BARCLAYS	21.604	0.286	3	U.K.
11	HSBC	21.682	0.127	4	U.K.
12	STANDARD CHARTERED	20.136	0.187	1	U.K.
13	BANK OF CHINA	21.200	0.160	1	China
14	ICBC	21.508	0.089	1	China
15	CHINA CON.BANK	21.281	0.092	1	China
16	BNP PARIBAS	21.684	0.136	3	France
17	CREDIT AGRICOLE	21.489	0.211	1	France
18	SOCIETE GENERALE	21.184	0.139	1	France
19	DEUTSCHE BANK	21.630	0.200	3	Germany
20	UNICREDIT	20.929	0.360	1	Italy
21	ING GROEP	21.156	0.103	1	Netherlands
22	SANTANDER	21.158	0.368	1	Spain
23	NORDEA BANK	20.476	0.326	1	Sweden
24	CREDIT SUISSE GROUP	20.744	0.339	2	Switzerland
25	UBS GROUP	21.008	0.251	1	Switzerland
26	MITSUBISHI UFJ	21.533	0.159	2	Japan
27	MIZUHO	21.247	0.233	1	Japan
28	SUMITOMO.MITSUI	21.044	0.125	1	Japan

* Debt ratio is defined as the ratio of total debt to total assets of a bank; and bank size is the log value of total assets; denominated in US dollars.

** Mean values during the sample period (2007-2015) are shown. The buckets assigned by BCBS correspond to required levels of additional common equity loss absorbency as percentage of risk-weighted assets from 3.5% (Bucket 5), 2.5%(Bucket 4), 2.0%(Bucket 3), 1.5%(Bucket 2) to 1%(Bucket 1)

Index	2007		2008		2009		2010		2011		2012		2013		2014	
	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99
1	-4.42	-18.56	-9.51	-24.28	-6.82	**-22.86**	-5.26	-15.80	-3.83	-15.73	-4.40	-19.31	-6.92	-21.43	-7.07	-21.35
2	-2.63	-16.45	-8.74	-24.88	-2.50	-16.56	-4.64	-16.19	-4.48	-15.36	-5.71	-18.30	-7.28	-21.57	-4.73	-18.25
3	-4.66	-17.31	-6.92	-18.47	-0.82	-11.82	-5.06	-14.95	-3.05	-14.84	-4.82	-18.19	-6.50	-20.15	-4.12	-14.58
4	-1.70	-12.93	-9.44	-22.82	-6.94	-21.58	-6.54	-15.46	-2.66	**-20.83**	-3.92	-18.81	-7.14	-18.17	-6.66	-21.95
5	-3.56	-16.39	-9.11	-21.69	-6.37	-19.18	-3.85	-17.05	-2.09	-14.33	-2.61	-14.33	-4.27	-17.90	-8.15	-20.78
6	-1.90	-15.42	-8.18	-24.63	-4.16	-15.08	-1.63	-13.27	-1.84	-16.46	**-6.20**	-18.91	-5.12	-19.62	**-9.04**	-21.90
7	-5.24	-18.12	-7.95	-23.87	-3.32	-15.84	-3.85	-16.84	-1.48	-13.49	-4.90	-19.43	-5.18	-19.22	-5.63	-16.99
8	-2.44	-17.56	-8.54	-23.64	-1.41	-14.21	-4.88	-19.73	-4.57	-17.58	-5.46	-19.58	**-6.37**	**-20.40**	-7.48	-19.21
9	-2.67	-14.61	**-10.50**	**-27.96**	-6.12	-21.23	-6.13	-17.59	-2.88	-14.35	-3.81	-16.71	-3.67	-15.90	-4.87	**-22.08**
10	-6.91	-16.61	-7.58	-20.66	-4.51	-20.03	-6.98	**-21.26**	-2.88	-14.00	-3.94	-16.57	-5.12	-15.73	-5.41	-16.76
11	-3.74	-16.19	-8.92	-26.06	-3.58	-14.61	-2.60	-14.38	-4.66	-18.93	-3.94	-14.63	-4.42	-20.19	-6.37	-19.03
12	-1.55	-16.17	-5.64	-22.48	-1.06	-15.25	-5.44	-19.85	-2.85	-16.33	-3.31	-16.80	-4.51	-18.77	-4.89	-17.87
13	-4.78	-19.19	-5.76	-18.34	-3.27	-16.17	-2.40	-14.00	-1.09	-12.58	-4.10	-14.90	-4.25	-18.35	-3.54	-18.56
14	-4.17	-16.66	-9.76	-24.92	0.01	-11.26	0.73	-11.78	-3.00	-16.90	-3.92	-14.35	-4.89	-17.61	-2.83	-17.14
15	-5.84	-17.40	-8.14	-24.49	-3.09	-14.92	-1.79	-15.24	-5.82	-16.46	-3.00	-13.39	-5.93	-19.09	-4.33	-17.41
16	-2.77	-12.30	-8.62	-23.98	-6.13	-17.26	**-8.17**	-19.27	**-6.51**	-19.84	-5.87	-18.56	-3.83	-12.70	-7.77	-19.33
17	-7.08	**-20.71**	-5.29	-20.80	-4.57	-17.93	-4.70	-15.85	-1.96	-14.47	-4.29	-18.55	-5.78	-18.38	-8.13	-19.76
18	-5.42	-17.52	-6.02	-21.67	-3.12	-15.35	-4.66	-15.05	-3.23	-11.01	-4.87	-15.76	-6.17	-20.22	-4.84	-21.29
19	-3.83	-15.56	-9.32	-21.81	-4.30	-18.00	-4.73	-16.62	-3.23	-19.01	-3.68	-16.32	-6.33	-17.29	-4.25	-18.99
20	-3.10	-13.11	-2.76	-15.27	-7.32	-20.27	-7.04	-19.76	-3.99	-15.27	-4.91	-17.71	-5.02	-17.28	-7.22	-20.50
21	-6.05	-16.85	-3.52	-18.06	-4.48	-19.08	-3.04	-14.19	-4.61	-16.01	-4.92	-18.90	-7.59	-18.27	-3.26	-17.94
22	-5.27	-15.35	-9.59	-25.05	-4.14	-16.56	-7.35	-20.51	-4.43	-15.48	-5.43	-18.64	-6.87	-15.35	-6.44	-18.10
23	**-9.22**	-19.27	-6.15	-19.45	-3.86	-16.75	-6.40	-20.20	-3.02	-13.87	-4.04	-15.74	-4.50	-17.58	-5.50	-17.16
24	-4.76	-16.43	-9.45	-21.13	**-8.13**	-20.55	-6.33	-18.58	-3.70	-15.78	-4.77	-15.16	-5.53	-19.72	-6.50	-21.21
25	-2.50	-13.09	-9.70	-25.01	-5.79	-16.17	-5.11	-16.91	-5.40	-14.16	-4.81	**-20.44**	-4.01	-17.11	-5.06	-16.56
26	-4.93	-16.59	-3.59	-16.86	-4.42	-15.28	-3.89	-17.65	-2.24	-14.12	-4.71	-18.78	-4.87	-17.16	-5.62	-15.66
27	-3.24	-15.41	-5.33	-16.49	-3.87	-19.50	-5.70	-16.20	-2.18	-12.94	-3.26	-15.45	-2.61	-14.40	-5.83	-18.06
28	-3.03	-14.68	-5.26	-14.21	-4.95	-15.60	-4.77	-17.85	-1.36	-13.99	-4.38	-13.69	-4.12	-15.66	-6.55	-20.17
Avg.NC	-5.44	-3.39	-14.40	-7.63	-9.72	-5.76	-7.99	-4.17	-8.20	-4.55	-8.27	-5.17	-3.77	-2.46	-4.40	-2.78

Table 3.2: Portfolio VaR in double-*t* copula model at 99% and 95% conditional on each node; in red: central nodes chosen by eigenvector centrality on the Pearson correlation matrix

Index	2007		2008		2009		2010		2011		2012		2013		2014	
	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99
1	-4.42	-18.56	-9.51	-24.28	-6.82	**-22.86**	-5.26	-15.80	-3.83	-15.73	-4.40	-19.31	-6.92	-21.43	-7.07	-21.35
2	-2.63	-16.45	-8.74	-24.88	-2.50	-16.56	-4.64	-16.19	-4.48	-15.36	-5.71	-18.30	-7.28	-21.57	-4.73	-18.25
3	-4.66	-17.31	-6.92	-18.47	-0.82	-11.82	-5.06	-14.95	-3.05	-14.84	-4.82	-18.19	-6.50	-20.15	-4.12	-14.58
4	-1.70	-12.93	-9.44	-22.82	-6.94	-21.58	-6.54	-15.46	-2.66	**-20.83**	-3.92	-18.81	-7.14	-18.17	-6.66	-21.95
5	-3.56	-16.39	-9.11	-21.69	-6.37	-19.18	-3.85	-17.05	-2.09	-14.33	-2.61	-14.33	-4.27	-17.90	-8.15	-20.78
6	-1.90	-15.42	-8.18	-24.63	-4.16	-15.08	-1.63	-13.27	-1.84	-16.46	**-6.20**	-18.91	-5.12	-19.62	**-9.04**	-21.90
7	-5.24	-18.12	-7.95	-23.87	-3.32	-15.84	-3.85	-16.84	-1.48	-13.49	-4.90	-19.43	-5.18	-19.22	-5.63	-16.99
8	-2.44	-17.56	-8.54	-23.64	-1.41	-14.21	-4.88	-19.73	-4.57	-17.58	-5.46	-19.58	**-6.37**	**-20.40**	-7.48	-19.21
9	-2.67	-14.61	**-10.50**	**-27.96**	-6.12	-21.23	-6.13	-17.59	-2.88	-14.35	-3.81	-16.71	-3.67	-15.90	-4.87	**-22.08**
10	-6.91	-16.61	-7.58	-20.66	-4.51	-20.03	-6.98	**-21.26**	-2.88	-14.00	-3.94	-16.57	-5.12	-15.73	-5.41	-16.76
11	-3.74	-16.19	-8.92	-26.06	-3.58	-14.61	-2.60	-14.38	-4.66	-18.93	-3.94	-14.63	-4.42	-20.19	-6.37	-19.03
12	-1.55	-16.17	-5.64	-22.48	-1.06	-15.25	-5.44	-19.85	-2.85	-16.33	-3.31	-16.80	-4.51	-18.77	-4.89	-17.87
13	-4.78	-19.19	-5.76	-18.34	-3.27	-16.17	-2.40	-14.00	-1.09	-12.58	-4.10	-14.90	-4.25	-18.35	-3.54	-18.56
14	-4.17	-16.66	-9.76	-24.92	0.01	-11.26	0.73	-11.78	-3.00	-16.90	-3.92	-14.35	-4.89	-17.61	-2.83	-17.14
15	-5.84	-17.40	-8.14	-24.49	-3.09	-14.92	-1.79	-15.24	-5.82	-16.46	-3.00	-13.39	-5.93	-19.09	-4.33	-17.41
16	-2.77	-12.30	-8.62	-23.98	-6.13	-17.26	**-8.17**	-19.27	**-6.51**	-19.84	-5.87	-18.56	-3.83	-12.70	-7.77	-19.33
17	-7.08	**-20.71**	-5.29	-20.80	-4.57	-17.93	-4.70	-15.85	-1.96	-14.47	-4.29	-18.55	-5.78	-18.38	-8.13	-19.76
18	-5.42	-17.52	-6.02	-21.67	-3.12	-15.35	-4.66	-15.05	-3.23	-11.01	-4.87	-15.76	-6.17	-20.22	-4.84	-21.29
19	-3.83	-15.56	-9.32	-21.81	-4.30	-18.00	-4.73	-16.62	-3.23	-19.01	-3.68	-16.32	-6.33	-17.29	-4.25	-18.99
20	-3.10	-13.11	-2.76	-15.27	-7.32	-20.27	-7.04	-19.76	-3.99	-15.27	-4.91	-17.71	-5.02	-17.28	-7.22	-20.50
21	-6.05	-16.85	-3.52	-18.06	-4.48	-19.08	-3.04	-14.19	-4.61	-16.01	-4.92	-18.90	-7.59	-18.27	-3.26	-17.94
22	-5.27	-15.35	-9.59	-25.05	-4.14	-16.56	-7.35	-20.51	-4.43	-15.48	-5.43	-18.64	-6.87	-15.35	-6.44	-18.10
23	**-9.22**	-19.27	-6.15	-19.45	-3.86	-16.75	-6.40	-20.20	-3.02	-13.87	-4.04	-15.74	-4.50	-17.58	-5.50	-17.16
24	-4.76	-16.43	-9.45	-21.13	**-8.13**	-20.55	-6.33	-18.58	-3.70	-15.78	-4.77	-15.16	-5.53	-19.72	-6.50	-21.21
25	-2.50	-13.09	-9.70	-25.01	-5.79	-16.17	-5.11	-16.91	-5.40	-14.16	-4.81	**-20.44**	-4.01	-17.11	-5.06	-16.56
26	-4.93	-16.59	-3.59	-16.86	-4.42	-15.28	-3.89	-17.65	-2.24	-14.12	-4.71	-18.78	-4.87	-17.16	-5.62	-15.66
27	-3.24	-15.41	-5.33	-16.49	-3.87	-19.50	-5.70	-16.20	-2.18	-12.94	-3.26	-15.45	-2.61	-14.40	-5.83	-18.06
28	-3.03	-14.68	-5.26	-14.21	-4.95	-15.60	-4.77	-17.85	-1.36	-13.99	-4.38	-13.69	-4.12	-15.66	-6.55	-20.17
Avg.NC	-5.44	-3.39	-14.40	-7.63	-9.72	-5.76	-7.99	-4.17	-8.20	-4.55	-8.27	-5.17	-3.77	-2.46	-4.40	-2.78

Table 3.3: Portfolio VaR in double-*t* copula model at 99% and 95% conditional on each node; in red: central nodes chosen by eigenvector centrality on the empirical tail dependence matrix

Index	2007		2008		2009		2010		2011		2012		2013		2014	
	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99	PVaR95	PVaR99
1	-4.42	-18.56	-9.51	-24.28	-6.82	**-22.86**	-5.26	-15.80	-3.83	-15.73	-4.40	-19.31	-6.92	-21.43	-7.07	-21.35
2	-2.63	-16.45	-8.74	-24.88	-2.50	-16.56	-4.64	-16.19	-4.48	-15.36	-5.71	-18.30	-7.28	-21.57	-4.73	-18.25
3	-4.66	-17.31	-6.92	-18.47	-0.82	-11.82	-5.06	-14.95	-3.05	-14.84	-4.82	-18.19	-6.50	-20.15	-4.12	-14.58
4	-1.70	-12.93	-9.44	-22.82	-6.94	-21.58	-6.54	-15.46	-2.66	**-20.83**	-3.92	-18.81	-7.14	-18.17	-6.66	-21.95
5	-3.56	-16.39	-9.11	-21.69	-6.37	-19.18	-3.85	-17.05	-2.09	-14.33	-2.61	-14.33	-4.27	-17.90	-8.15	-20.78
6	-1.90	-15.42	-8.18	-24.63	-4.16	-15.08	-1.63	-13.27	-1.84	-16.46	**-6.20**	-18.91	-5.12	-19.62	**-9.04**	-21.90
7	-5.24	-18.12	-7.95	-23.87	-3.32	-15.84	-3.85	-16.84	-1.48	-13.49	-4.90	-19.43	-5.18	-19.22	-5.63	-16.99
8	-2.44	-17.56	-8.54	-23.64	-1.41	-14.21	-4.88	-19.73	-4.57	-17.58	-5.46	-19.58	**-6.37**	**-20.40**	-7.48	-19.21
9	-2.67	-14.61	**-10.50**	**-27.96**	-6.12	-21.23	-6.13	-17.59	-2.88	-14.35	-3.81	-16.71	-3.67	-15.90	-4.87	**-22.08**
10	-6.91	-16.61	-7.58	-20.66	-4.51	-20.03	-6.98	**-21.26**	-2.88	-14.00	-3.94	-16.57	-5.12	-15.73	-5.41	-16.76
11	-3.74	-16.19	-8.92	-26.06	-3.58	-14.61	-2.60	-14.38	-4.66	-18.93	-3.94	-14.63	-4.42	-20.19	-6.37	-19.03
12	-1.55	-16.17	-5.64	-22.48	-1.06	-15.25	-5.44	-19.85	-2.85	-16.33	-3.31	-16.80	-4.51	-18.77	-4.89	-17.87
13	-4.78	-19.19	-5.76	-18.34	-3.27	-16.17	-2.40	-14.00	-1.09	-12.58	-4.10	-14.90	-4.25	-18.35	-3.54	-18.56
14	-4.17	-16.66	-9.76	-24.92	0.01	-11.26	0.73	-11.78	-3.00	-16.90	-3.92	-14.35	-4.89	-17.61	-2.83	-17.14
15	-5.84	-17.40	-8.14	-24.49	-3.09	-14.92	-1.79	-15.24	-5.82	-16.46	-3.00	-13.39	-5.93	-19.09	-4.33	-17.41
16	-2.77	-12.30	-8.62	-23.98	-6.13	-17.26	**-8.17**	-19.27	**-6.51**	-19.84	-5.87	-18.56	-3.83	-12.70	-7.77	-19.33
17	-7.08	**-20.71**	-5.29	-20.80	-4.57	-17.93	-4.70	-15.85	-1.96	-14.47	-4.29	-18.55	-5.78	-18.38	-8.13	-19.76
18	-5.42	-17.52	-6.02	-21.67	-3.12	-15.35	-4.66	-15.05	-3.23	-11.01	-4.87	-15.76	-6.17	-20.22	-4.84	-21.29
19	-3.83	-15.56	-9.32	-21.81	-4.30	-18.00	-4.73	-16.62	-3.23	-19.01	-3.68	-16.32	-6.33	-17.29	-4.25	-18.99
20	-3.10	-13.11	-2.76	-15.27	-7.32	-20.27	-7.04	-19.76	-3.99	-15.27	-4.91	-17.71	-5.02	-17.28	-7.22	-20.50
21	-6.05	-16.85	-3.52	-18.06	-4.48	-19.08	-3.04	-14.19	-4.61	-16.01	-4.92	-18.90	-7.59	-18.27	-3.26	-17.94
22	-5.27	-15.35	-9.59	-25.05	-4.14	-16.56	-7.35	-20.51	-4.43	-15.48	-5.43	-18.64	-6.87	-15.35	-6.44	-18.10
23	**-9.22**	-19.27	-6.15	-19.45	-3.86	-16.75	-6.40	-20.20	-3.02	-13.87	-4.04	-15.74	-4.50	-17.58	-5.50	-17.16
24	-4.76	-16.43	-9.45	-21.13	**-8.13**	-20.55	-6.33	-18.58	-3.70	-15.78	-4.77	-15.16	-5.53	-19.72	-6.50	-21.21
25	-2.50	-13.09	-9.70	-25.01	-5.79	-16.17	-5.11	-16.91	-5.40	-14.16	-4.81	**-20.44**	-4.01	-17.11	-5.06	-16.56
26	-4.93	-16.59	-3.59	-16.86	-4.42	-15.28	-3.89	-17.65	-2.24	-14.12	-4.71	-18.78	-4.87	-17.16	-5.62	-15.66
27	-3.24	-15.41	-5.33	-16.49	-3.87	-19.50	-5.70	-16.20	-2.18	-12.94	-3.26	-15.45	-2.61	-14.40	-5.83	-18.06
28	-3.03	-14.68	-5.26	-14.21	-4.95	-15.60	-4.77	-17.85	-1.36	-13.99	-4.38	-13.69	-4.12	-15.66	-6.55	-20.17
Avg.NC	-5.44	-3.39	-14.40	-7.63	-9.72	-5.76	-7.99	-4.17	-8.20	-4.55	-8.27	-5.17	-3.77	-2.46	-4.40	-2.78

Table 3.4: Portfolio VaR in double-*t* copula model at 99% and 95% conditional on each node; in red: central nodes chosen by tail matrix norm measure

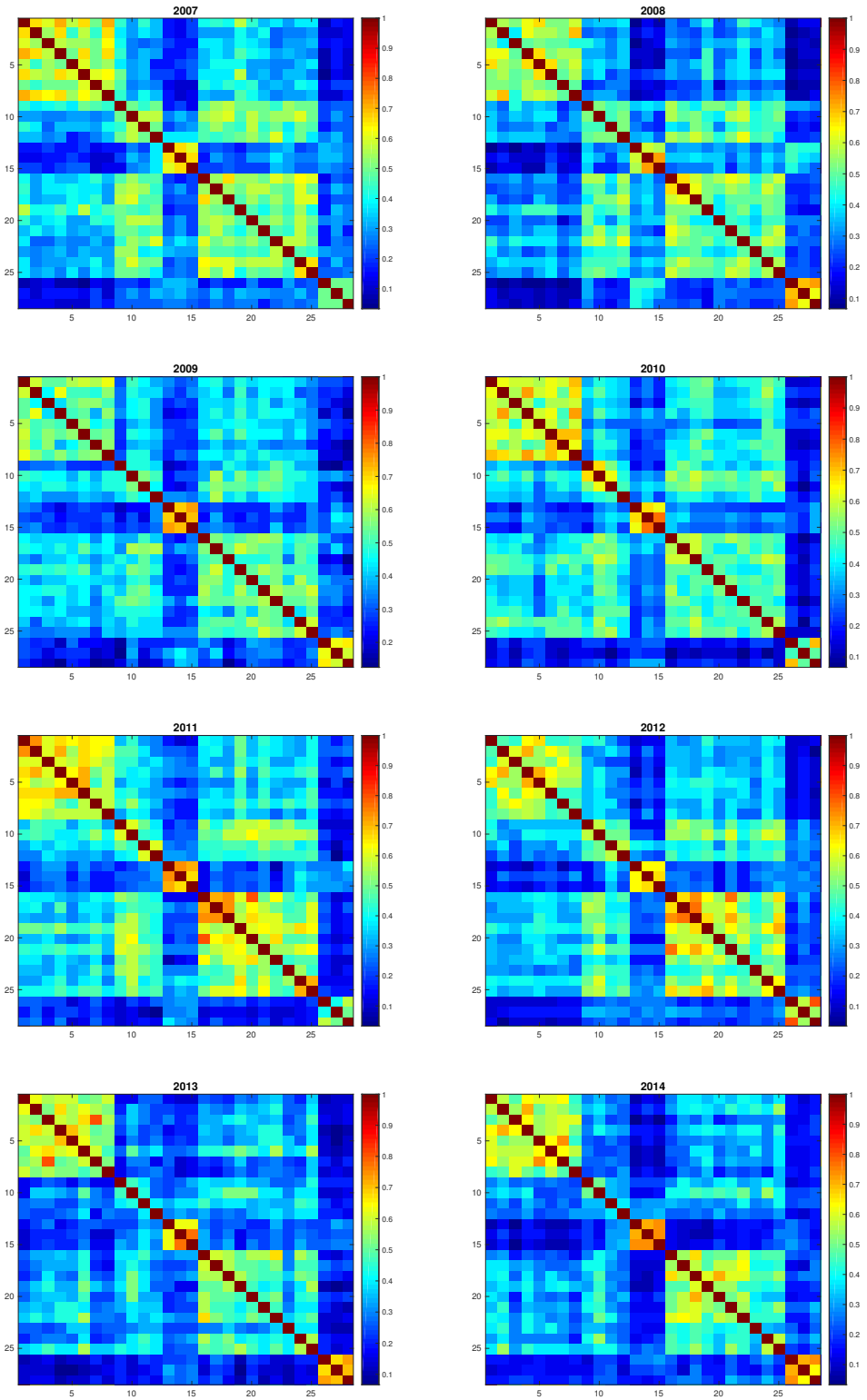

Figure 3.1: Empirical tail dependence matrices for 28 SIFIs

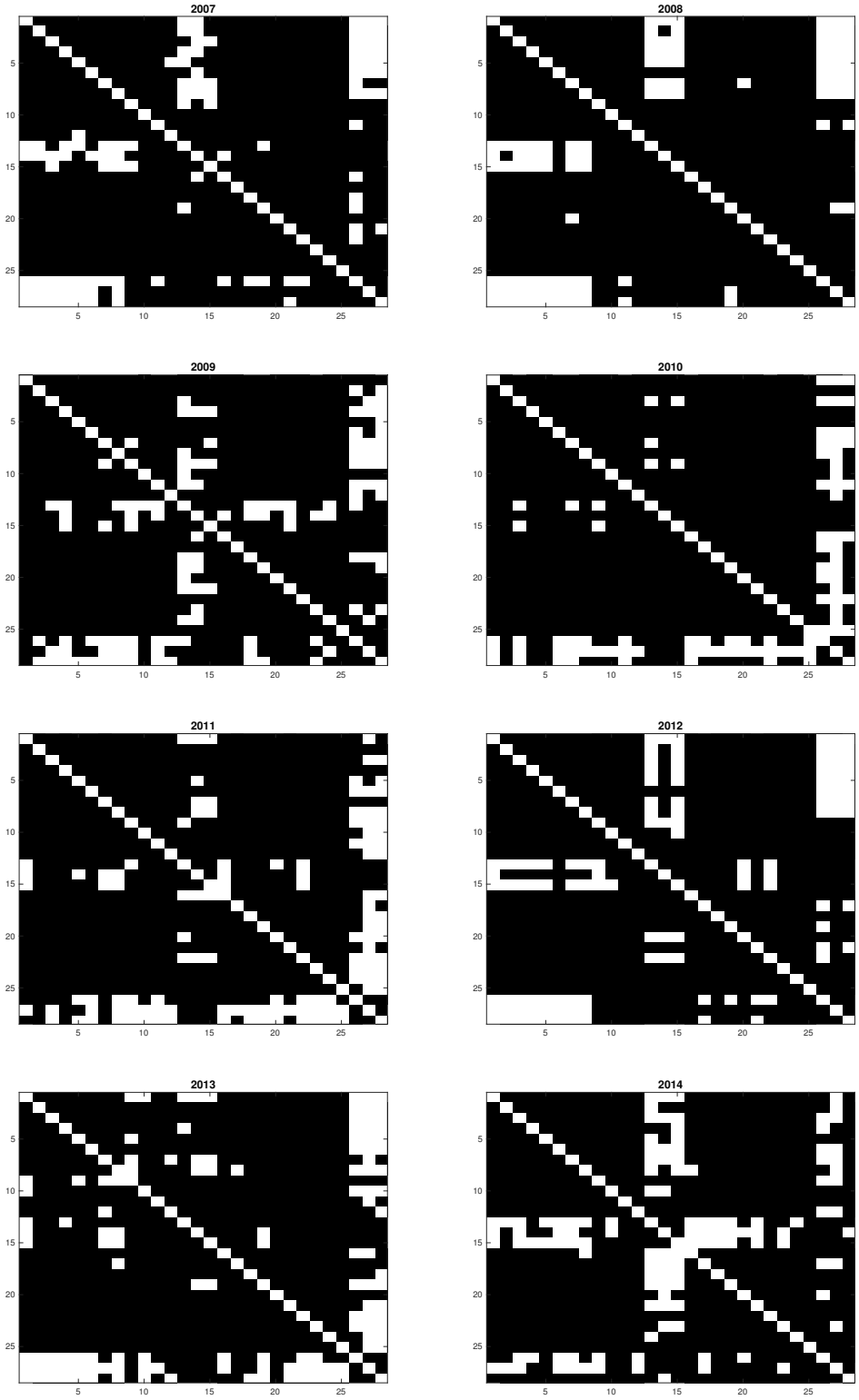

Figure 3.2: Binary adjacency matrices for 28 SIFIs obtained from empirical tail dependence in Figure 3.1 (black: 1, white: 0)

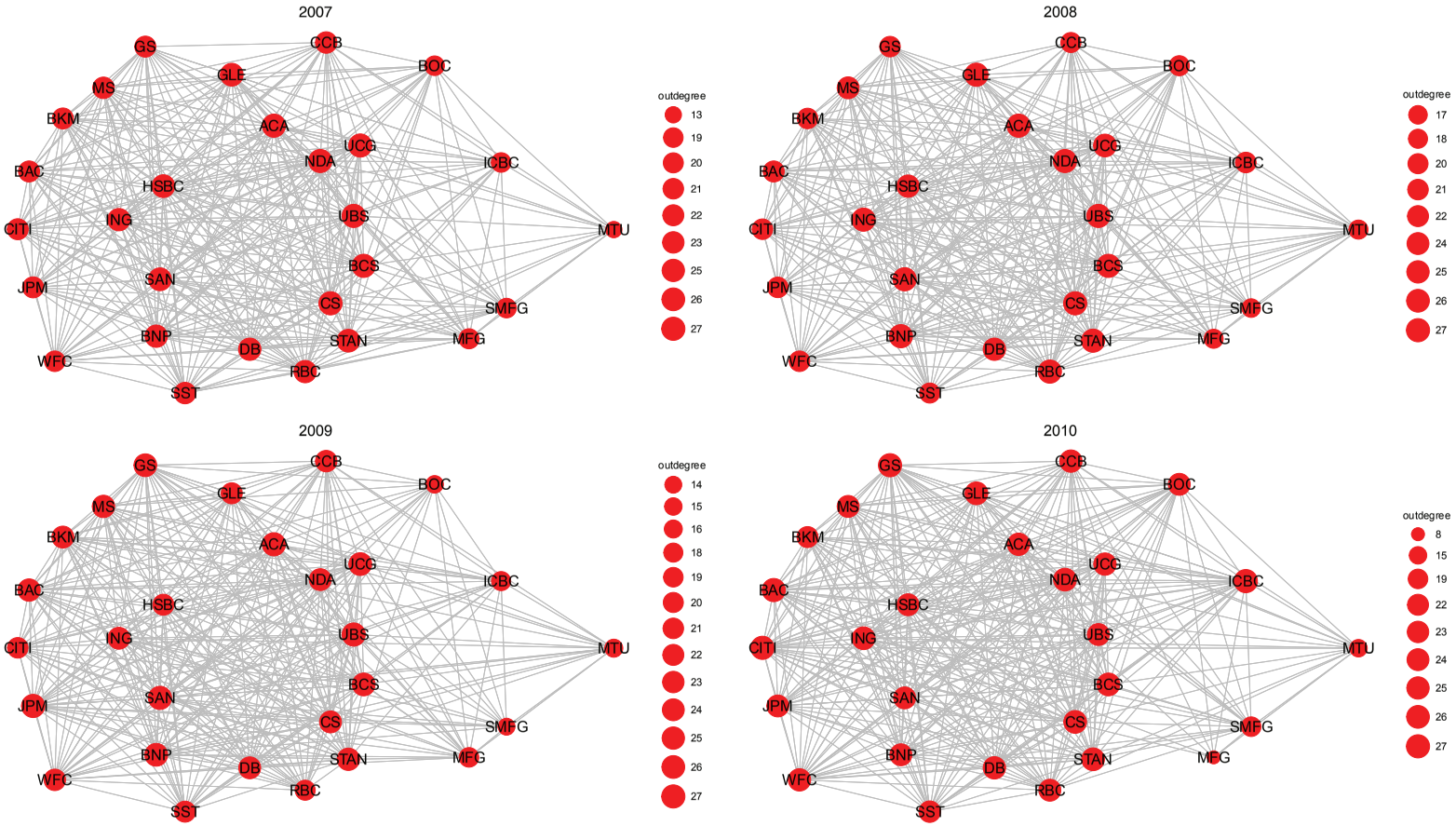

Figure 3.3: SIFI network structures produced by adjacency analysis on the empirical tail dependence matrix

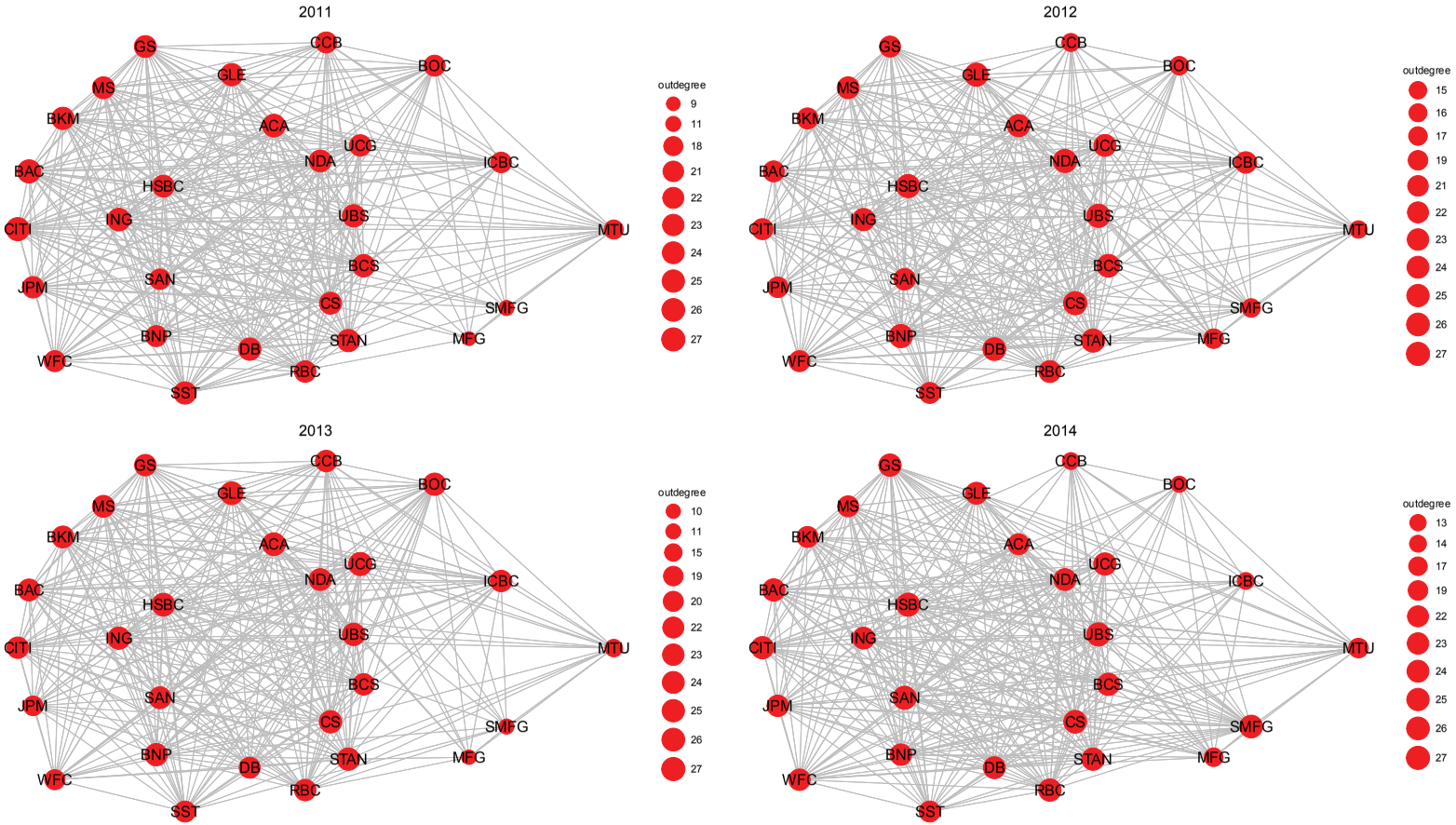

Figure 3.4: SIFI network structures produced by adjacency analysis on the empirical tail dependence matrix

Chapter 4

Appendix

4.1 Non-positive Lasso quantile regression optimization problem

The Lasso-penalized QR problem with an additional non-positivity constraint takes the following form:

$$\begin{aligned} \underset{(\xi,\zeta,\eta,\tilde{\beta})\in\mathbb{R}^{2n}_{+}\times\mathbb{R}^{2p}}{\text{minimize}} \quad & \tau 1_n^\top \xi + (1-\tau) 1_n^\top \zeta + \lambda 1_n^\top \eta \\ \text{subject to} \quad & \xi - \zeta = Y + X\tilde{\beta}, \\ & \xi \geq 0, \\ & \zeta \geq 0, \\ & \eta \geq \tilde{\beta}, \\ & \eta \geq -\tilde{\beta}, \\ & \tilde{\beta} \geq 0, \quad \tilde{\beta} \stackrel{\text{def}}{=} -\beta \end{aligned} \tag{4.1}$$

with "slack" variables ξ, ζ and η.

Transformed into matrix form, this problem can be equivalently re-written as

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax = b, \;\; Bx \leq 0, \\ & l_i \leq x_i \leq u_i, \;\; i \in \mathscr{I} \\ & 0 \leq x_i, \;\; i \in \mathscr{J} \end{aligned}$$

where $\mathscr{I}$, $\mathscr{J}$ are disjoint index sets such that $\mathscr{I} \cup \mathscr{J} = \{1,2,\ldots,n\}$, $\mathscr{I} \cap \mathscr{J} = \emptyset$,

$A = \begin{pmatrix} I_n & -I_n & 0 & X \end{pmatrix}, b = Y, x = \begin{pmatrix} \xi & \zeta & \eta & \beta \end{pmatrix}^\top$, and

$$c = \begin{pmatrix} \tau 1_n \\ (1-\tau)1_n \\ \lambda 1_p \\ 01_p \end{pmatrix}, \quad B = \begin{pmatrix} -E_{p\times n} & 0 & 0 & 0 \\ 0 & -E_{p\times n} & 0 & 0 \\ 0 & 0 & -I_p & I_p \\ 0 & 0 & -I_p & -I_p \\ 0 & 0 & 0 & I_p \end{pmatrix}$$

where I_p is $p \times p$ identity matrix; $E_{p\times n} = \begin{pmatrix} I_n \\ 0 \end{pmatrix}$. Denote the dimension of matrix A as $m \times \tilde{n}$. Without loss of generality, as noted by Zhang (1998), it is also assumed that for some positive integer $n_u \leq \tilde{n}$

$$\mathscr{J} = \{1, 2, \ldots, n_u\}, \quad \mathscr{J} = \{n_u + 1, n_u + 2, \ldots, \tilde{n}\}.$$

A suitable algorithm to compute the solution is an efficient large-scale primal-dual infeasible-interior-point algorithm using the Newton method as solver, outlined in Zhang (1998).

4.2 Proof of Proposition 1.1.3.1

Define the Cornish-Fisher expansion:

$$q_\alpha \stackrel{\text{def}}{=} z_\alpha + (z_\alpha^2 - 1)s + (z_\alpha^3 - 3z_\alpha)k - (2z_\alpha^3 - 5z_\alpha)s^2,$$

where $s \stackrel{\text{def}}{=} S/6$, $k \stackrel{\text{def}}{=} K/24$, S and K are skewness and excess kurtosis, respectively; $z_\alpha \stackrel{\text{def}}{=} \Phi^{-1}(\alpha)$.

Re-write:

$$q_\alpha = a_0 + a_1 z_\alpha + a_2 z_\alpha^2 + a_3 z_\alpha^3, \tag{4.2}$$

where $a_0 = -s$, $a_1 = 1 - 3k + 5s^2$, $a_2 = s$, $a_3 = k - 2s^2$.

Define the *conditional Value-at-Risk* or *expected shortfall* as the average of the worst $100(1-\alpha)\%$ of losses:

$$\text{CVaR}_\alpha \stackrel{\text{def}}{=} \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_q \, \mathrm{d}\, q,$$

where $\text{VaR}_q \stackrel{\text{def}}{=} -\Phi^{-1}(q)\sigma\sqrt{T}$, σ is volatility.

Observe:

$$\text{CVaR}_\alpha = -\frac{1}{1-\alpha}\int_\alpha^1 \Phi^{-1}(\alpha)\sigma\sqrt{T}\,\mathrm{d}q \tag{4.3}$$

$$= -\frac{\sigma\sqrt{T}}{1-\alpha}\int_\alpha^1 \Phi^{-1}(q)\,\mathrm{d}q \tag{4.4}$$

$$= -\frac{\sigma\sqrt{T}}{1-\alpha}\int_{-\infty}^{-z_\alpha} u\varphi(u)\,\mathrm{d}u, \tag{4.5}$$

where (4.5) follows from the change of variable: $u = z_q = \Phi^{-1}(q)$.

Substitute (4.2) into (4.5):

$$\begin{aligned}\text{CVaR}_\alpha &= -\frac{\sigma\sqrt{T}}{1-\alpha}\int_{-\infty}^{-z_\alpha}\left(a_0 + a_1 z + a_2 z^2 + a_3 z^3\right)\varphi(z)\,\mathrm{d}z \\ &= a_0 A_0 + a_1 A_1 + a_2 A_2 + a_3 A_3,\end{aligned}$$

$$\begin{aligned}A_0 &= -\frac{\sigma\sqrt{T}}{1-\alpha}\int_{-\infty}^{-z_\alpha}\varphi(z)\,\mathrm{d}z = -\sigma\sqrt{T}, \\ A_1 &= -\frac{\sigma\sqrt{T}}{1-\alpha}\int_{-\infty}^{-z_\alpha} z\varphi(z)\,\mathrm{d}z = \frac{\sigma\sqrt{T}}{1-\alpha}\varphi(z_\alpha), \\ A_2 &= -\frac{\sigma\sqrt{T}}{1-\alpha}\int_{-\infty}^{-z_\alpha} z^2\varphi(z)\,\mathrm{d}z = -\sigma\sqrt{T}\left(\frac{\varphi(z_\alpha)z_\alpha}{1-\alpha}+1\right), \\ A_3 &= -\frac{\sigma\sqrt{T}}{1-\alpha}\int_{-\infty}^{-z_\alpha} z^3\varphi(z)\,\mathrm{d}z = \frac{\sigma\sqrt{T}}{1-\alpha}(z_\alpha^2+2)\varphi(z_\alpha).\end{aligned}$$

Collecting terms, simplifying and replacing σ, S and K with $\sigma_{P,t}(w_t)$, $S_{P,t}(w_t)$ and $K_{P,t}(w_t)$, respectively, yields the result.

4.3 Accuracy Criteria

1. Standardized L_2-norm
$$\text{Dev} \stackrel{\text{def}}{=} \frac{\|\beta - \hat{\beta}\|_2}{\|\beta\|_2}$$
2. Sign consistency
$$\text{Acc} \stackrel{\text{def}}{=} \sum_{j=1}^{p} |\operatorname{sign}(\beta_j) - \operatorname{sign}(\hat{\beta}_j)|$$
3. Least angle
$$\text{Angle} \stackrel{\text{def}}{=} \frac{<\beta, \hat{\beta}>}{\|\beta\|_2 \cdot \|\hat{\beta}\|_2}$$

4. Estimate of true model dimension:

$$\text{Est} \stackrel{\text{def}}{=} \hat{q}$$

5. Empirical risk

$$\text{Risk} \stackrel{\text{def}}{=} \sqrt{n^{-1} \sum_{i=1}^{n} \left\{ X_i^\top (\beta - \hat{\beta}) \right\}^2}$$

4.4 Moneyness scaling formula

Given the asset (S&P 500) price dynamics

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t^{\mathsf{Q}} \tag{4.6}$$

with interest rate r and volatility σ; W_t^{Q} standard Brownian motion under the risk-neutral measure Q, the (L)ETF dynamics is given by:

$$\begin{aligned} \frac{dL_t}{L_t} &= \beta \left(\frac{dS_t}{S_t} \right) - \{(\beta - 1)r + c\}dt \\ &= (r - c)dt + \beta \sigma dW_t^{\mathsf{Q}}, \end{aligned} \tag{4.7}$$

where $0 \leq c \ll r$ is the (L)ETF expense ratio (approximates an annual fee charged by the ETF from the shareholders to cover the fund's operating expenses). Then the general solution of (4.7) is given by:

$$L_T = L_t \exp\left\{ (r - c)(T - t) - \frac{\beta^2}{2} \int_t^T \sigma_s^2 ds + \beta \int_t^T \sigma_s dW_s^{\mathsf{Q}} \right\}. \tag{4.8}$$

If we write (4.8) for $L_T^{(\beta_1)}$, $L_T^{(\beta_2)}$, we obtain

$$\frac{L_T^{(\beta_1)}}{e^{(r-c)\tau} L_t^{(\beta_1)}} = \exp\left(-\frac{\beta_1^2}{2} \int_0^\tau \sigma_s^2 ds + \beta_1 \int_0^\tau \sigma_s dW_s^{\mathsf{Q}} \right), \tag{4.9}$$

$$\frac{L_T^{(\beta_2)}}{e^{(r-c)\tau} L_t^{(\beta_2)}} = \exp\left(-\frac{\beta_2^2}{2} \int_0^\tau \sigma_s^2 ds + \beta_2 \int_0^\tau \sigma_s dW_s^{\mathsf{Q}} \right), \tag{4.10}$$

where σ_s is the instantaneous volatility at time s.

From (4.10) it follows:

$$\int_0^\tau \sigma_s dW_s^{\mathsf{Q}} = \frac{\log\left(\frac{L_T^{(\beta_2)}}{e^{(r-c)\tau}L_t^{(\beta_2)}}\right) + \frac{\beta_2^2}{2}\int_0^\tau \sigma_s^2 ds}{\beta_2}. \tag{4.11}$$

Substitute (4.11) into (4.9) to eliminate the stochastic term $\int_0^\tau \sigma_s dW_s^{\mathsf{Q}}$ and obtain:

$$\frac{L_T^{(\beta_1)}}{e^{(r-c)\tau}L_t^{(\beta_1)}} = \exp\left\{-\frac{\beta_1}{2}(\beta_1-\beta_2)\int_0^\tau \sigma_s^2 \mathrm{d}s\right\}\left\{\frac{L_T^{(\beta_2)}}{e^{(r-c)\tau}L_t^{(\beta_2)}}\right\}^{\frac{\beta_1}{\beta_2}} \tag{4.12}$$

Now take logarithms and expectations conditioned on $K^{(\beta_1)} = L_T^{(\beta_1)}$ and $K^{(\beta_2)} = L_T^{(\beta_2)}$ and obtain:

$$\begin{aligned}\log(k_f{}^{(\beta_1)}) = &-\frac{\beta_1}{2}(\beta_1-\beta_2)\mathsf{E}^{\mathsf{Q}}\left(\int_0^\tau \sigma_s^2 \mathrm{d}s \,\middle|\, K^{(\beta_1)} = L_T^{(\beta_1)}, K^{(\beta_2)} = L_T^{(\beta_2)}\right)\\ &+\frac{\beta_1}{\beta_2}\log(k_f{}^{(\beta_2)})\end{aligned}$$

Assuming constant σ and exponentiating, one obtains (2.6).

4.5 The local linear M-smoothing estimator

M-type smoothers apply a nonquadratic loss function $\rho(\cdot)$ to make estimation more robust. Given the model

$$Y_i = m(X_i) + \varepsilon_i, \tag{4.13}$$

where $Y_i \in \mathbb{R}$, $X_i \in \mathbb{R}^d$, $\varepsilon_i \stackrel{\text{def}}{=} \sigma(X_i)u_i$, $u_i \sim (0,1)$, iid, $\mathscr{X} \stackrel{\text{def}}{=} \{(X_i, Y_i); 1 \le i \le n\}$, the local linear M-smoothing estimator is obtained from:

$$\min_{\alpha\in\mathbb{R},\beta\in\mathbb{R}^p} \sum_{i=1}^n \rho\left\{Y_i - \alpha - \beta^\top(X_i - x)\right\} W_{ih}(x), \tag{4.14}$$

where

$$W_{hi}(x) \stackrel{\text{def}}{=} \frac{h^{-2}K'\{(x-X_i)/h\}}{\widehat{f}_h(x)} - \frac{K_h(x-X_i)\widehat{f}_h'(x)}{\widehat{f}_h^2(x)} \tag{4.15}$$

is a kernel weight sequence with $\widehat{f}_h'(x) \stackrel{\text{def}}{=} n^{-1}\sum_{i=1}^n K_h'(x-X_i)$, h is the bandwidth, K is a kernel function; $\int K(u)du = 1$, $K_h(\cdot) \stackrel{\text{def}}{=} h^{-1}K(\cdot/h)$. The function $\rho(\cdot)$ is designed to provide more robustness than the quadratic loss. An example of such a function is given by Huber (1964), see also Härdle (1989):

$$\rho(u) = \begin{cases} 0.5u^2, & \text{if } |u| \le c; \\ c|u| - 0.5c^2 & \text{if } |u| > c. \end{cases}, \tag{4.16}$$

with the constant c regulating the degree of resistance.

4.6 Proof of Proposition 2.3.1.1

The proof is based on the usage of the Dupire formula, which is a result from local volatility analysis, see **?**. It is known that the Dupire formula allows to compute the local volatility of a European option, defined by

$$\sigma^2_{K,T}(S_t,t) \overset{\text{def}}{=} \mathsf{E}^{\mathsf{Q}}\{\sigma^2(S_T,T,t,\cdot)|S_T = K,\mathscr{F}_t\} \tag{4.17}$$

$$= \frac{\frac{\partial C}{\partial T}}{\frac{K^2}{2}\frac{\partial^2 C}{\partial K^2}} \tag{4.18}$$

Then, given the Heston setup, $\sigma^2(S_T,T,t,\cdot) = V_t$, some constant $LM^{(1)}$, it follows that

$$\begin{aligned}
\mathsf{E}^{\mathsf{Q}}\left(\int_0^T V_t \mathrm{d}t \,\middle|\, \log\left(\frac{S_T}{S_0}\right) = LM^{(1)}\right) &= \int_0^T \mathsf{E}^{\mathsf{Q}}\left(V_t \,\middle|\, \log\left(\frac{S_T}{S_0}\right) = LM^{(1)}\right)\mathrm{d}t \\
&= \int_0^T \mathsf{E}^{\mathsf{Q}}\left(V_t \,\middle|\, S_T = S_0\exp(LM^{(1)})\right)\mathrm{d}t \\
&= \int_0^T \mathsf{E}^{\mathsf{Q}}\left(V_t \,\middle|\, S_T = \widetilde{K}\right)\mathrm{d}t,
\end{aligned}$$

where $\widetilde{K}$ is some constant strike price for a european call option on the ETF X. Then applying the Dupire formula (4.18) to $\mathsf{E}^{\mathsf{Q}}\left(V_t \,\middle|\, S_T = \widetilde{K}\right)$ yields

$$\begin{aligned}
\mathsf{E}^{\mathsf{Q}}\left(\int_0^T V_t \mathrm{d}t \,\middle|\, \log\left(\frac{S_T}{S_0}\right) = LM^{(1)}\right) &= \int_0^T \mathsf{E}^{\mathsf{Q}}\left(V_t \,\middle|\, S_T = \widetilde{K}\right)\mathrm{d}t \\
&= \frac{\frac{\partial C_H}{\partial T}}{\frac{\widetilde{K}^2}{2}\frac{\partial^2 C_H}{\partial \widetilde{K}^2}},
\end{aligned}$$

where C_H is the Heston price of the call option. The partial derivatives $\partial C_H/\partial T$ and $\partial^2 C_H/\partial \widetilde{K}^2$ are the "Heston Greeks" and are given, e.g., in Rouah (2013). After simplifications, one obtains (2.18).

4.7 Conditional pair Gaussian copula

The expression (3.15) is derived noting that $C_{U_j|V}(u_j|v) = \partial C_{U_j,V}(u_j,v)/\partial v$; denoting $\Phi_2(x,y;\rho)$ a bivariate cumulative distribution function with correlation ρ, it

follows then

$$C_{U_j|V}(u_j|v) = \frac{\partial C_{U_j,V}(u_j,v)}{\partial v} \tag{4.19}$$

$$= \frac{\partial \Phi_2\left(\Phi^{-1}(u_j),\Phi^{-1}(v);\alpha_{j1}\right)}{\partial v} \tag{4.20}$$

$$= \frac{\partial \Phi_2\left(\Phi^{-1}(u_j),\Phi^{-1}(v);\alpha_{j1}\right)}{\partial \Phi^{-1}(v)} \frac{\partial \Phi^{-1}(v)}{\partial v} \tag{4.21}$$

$$= \int_{-\infty}^{\Phi^{-1}(u_j)} \varphi_2\left(x,\Phi^{-1}(v);\alpha_{j1}\right) \frac{1}{\varphi(\Phi^{-1}(v))} \mathrm{d}x \tag{4.22}$$

$$= \int_{-\infty}^{\Phi^{-1}(u_j)} \frac{1}{\sqrt{2\pi(1-\alpha_{j1}^2)}} \exp\left\{-\frac{(x-\alpha_{j1}\Phi^{-1}(v))^2}{2(1-\alpha_{j1}^2)}\right\} \mathrm{d}x \tag{4.23}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\Phi^{-1}(u_j)-\alpha_{j1}\Phi^{-1}(v)}{\sqrt{1-\alpha_{j1}^2}}} \exp\left\{-\frac{u^2}{2}\right\} \mathrm{d}u \tag{4.24}$$

$$= \Phi\left\{\frac{\Phi^{-1}(u_j)-\alpha_{j1}\Phi^{-1}(v)}{\sqrt{1-\alpha_{j1}^2}}\right\}, \tag{4.25}$$

where the sixth equality comes from integration by substitution. The resulting expression (3.15) also can be obtained from the 1-factor correlation structure (3.13).

4.8 General conditional pair copula

Assume that (3.23) holds and that Z_i, ε_j are independent, identically distributed variables with cumulative distribution functions F_{Z_i}, F_{ε_j}. Then $U_j \stackrel{\text{def}}{=} F_{Z_j}(Z_j)$, $j = 1,\ldots,n$ are conditionally independent given $Z_i = z_i$. Moreover,

$$C_{U_j|Z_i}(u_j|z_i) = F_{U_j|Z_i}(u_j|z_i) \tag{4.26}$$

$$= \mathbb{P}\left(F_{Z_j}(Z_j) \le u_j\right) \tag{4.27}$$

$$= \mathbb{P}\left(Z_j \le F_{Z_j}^{-1}(u_j)\right) \tag{4.28}$$

$$= \mathbb{P}\left(\alpha_{j|i}Z_i + \sqrt{1-\alpha_{j|i}^2}\varepsilon_j \le F_{Z_j}^{-1}(u_j)\right) \tag{4.29}$$

$$= \mathbb{P}\left(\varepsilon_j \le \frac{F_{Z_j}^{-1}(u_j)-\alpha_{j|i}z_i}{\sqrt{1-\alpha_{j|i}^2}}\right) \tag{4.30}$$

$$= F_{\varepsilon_j}\left(\frac{F_{Z_j}^{-1}(u_j)-\alpha_{j|i}z_i}{\sqrt{1-\alpha_{j|i}^2}}\right), \tag{4.31}$$

And we obtain expression (3.24).

4.9 Proof of Proposition 3.3.3.1

According to the properties of functions with regular variation, see Feller (1971), given that the tails of two variables W, ε_j are different but symmetric, then $\mathsf{P}(W+\varepsilon_j<-s)=s^{-\alpha}(A_W+A_\varepsilon)+\mathscr{o}(s^{-\alpha})$, see also Hyung and de Vries (2007), Oh and Patton (2015). Then it follows:

$$\begin{aligned}\mathsf{P}(Z_j<-s) &= \mathsf{P}(\theta_{j|1}W+\sqrt{1-\theta_{j|1}^2}\varepsilon_j<-s)\\ &= \mathsf{P}(\theta_{j|1}W<-s)+\mathsf{P}(\sqrt{1-\theta_{j|1}^2}\varepsilon_j<-s)+\mathscr{o}(s^{-\alpha})\\ &= A_W\left(\frac{s}{\theta_{j|1}}\right)^{-\alpha}+A_\varepsilon\left(\frac{s}{\sqrt{1-\theta_{j|1}^2}}\right)^{-\alpha}\\ &= s^{-\alpha}\left(A_W\theta_{j|1}^\alpha+A_\varepsilon(1-\theta_{j|1}^2)^{\alpha/2}\right), \end{aligned} \tag{4.32}$$

as $s\to\infty$.

Consider two different dynamics of Z_i and Z_j, $\theta_{i|1}\neq\theta_{j|1}$. Then, following Oh and Patton (2015), we find the link between two thresholds s_i and s_j. The relation between $s_i/\theta_{i|1}$ and $s_j/\theta_{j|1}$ depends if the value of the expression

$$A_W\theta_{j|1}^\alpha\theta_{i|1}^\alpha+A_\varepsilon(1-\theta_{j|1}^2)^{\alpha/2}\theta_{i|1}^\alpha \tag{4.33}$$

is smaller or larger compared to the value of

$$A_W\theta_{i|1}^\alpha\theta_{j|1}^\alpha+A_\varepsilon(1-\theta_{i|1}^2)^{\alpha/2}\theta_{j|1}^\alpha. \tag{4.34}$$

This follows from the observation that

$$\begin{aligned}\mathsf{P}(Z_i<-s_i,Z_j<-s_j) &\approx \mathsf{P}(\theta_{i|1}W<-s_i,\theta_{j|1}W<-s_j)+\mathscr{o}(s^{-\alpha})\\ &\approx \mathsf{P}\left(W<-\frac{s_i}{\theta_{i|1}},W<-\frac{s_j}{\theta_{j|1}}\right)\\ &\approx \mathsf{P}\left(W<\min\left\{-\frac{s_i}{\theta_{i|1}},-\frac{s_j}{\theta_{j|1}}\right\}\right), \end{aligned} \tag{4.35}$$

as $s\to\infty$.

Furthermore,

$$\mathsf{P}\left(W<\min\left\{-s_i/\theta_{i|1},-s_j/\theta_{j|1}\right\}\right)=\begin{cases} s_i^{-\alpha}A_W\theta_{i|1}^\alpha & \text{if } |s_i/\theta_{i|1}|>|s_j/\theta_{j|1}|\,,\\ s_j^{-\alpha}A_W\theta_{j|1}^\alpha & \text{if } |s_i/\theta_{i|1}|<|s_j/\theta_{j|1}|\,. \end{cases} \tag{4.36}$$

The condition $|s_i/\theta_{i|1}| > |s_j/\theta_{j|1}|$ is fulfilled when simultaneously

$$A_W\theta_{i|1}^{\alpha}\theta_{j|1}^{\alpha} + A_\varepsilon(1-\theta_{i|1}^2)^{\alpha/2}\theta_{j|1}^{\alpha} > A_W\theta_{j|1}^{\alpha}\theta_{i|1}^{\alpha} + A_\varepsilon(1-\theta_{j|1}^2)^{\alpha/2}\theta_{i|1}^{\alpha}, \tag{4.37}$$

$$\theta_{i|1} < \theta_{j|1}, \tag{4.38}$$

or

$$A_W\theta_{i|1}^{\alpha}\theta_{j|1}^{\alpha} + A_\varepsilon(1-\theta_{i|1}^2)^{\alpha/2}\theta_{j|1}^{\alpha} < A_W\theta_{j|1}^{\alpha}\theta_{i|1}^{\alpha} + A_\varepsilon(1-\theta_{j|1}^2)^{\alpha/2}\theta_{i|1}^{\alpha}, \tag{4.39}$$

$$\theta_{i|1} > \theta_{j|1}. \tag{4.40}$$

On the other hand, the condition $|s_i/\theta_{i|1}| < |s_j/\theta_{j|1}|$ is fulfilled when simultaneously

$$A_W\theta_{i|1}^{\alpha}\theta_{j|1}^{\alpha} + A_\varepsilon(1-\theta_{i|1}^2)^{\alpha/2}\theta_{j|1}^{\alpha} > A_W\theta_{j|1}^{\alpha}\theta_{i|1}^{\alpha} + A_\varepsilon(1-\theta_{j|1}^2)^{\alpha/2}\theta_{i|1}^{\alpha}, \tag{4.41}$$

$$\theta_{i|1} > \theta_{j|1}, \tag{4.42}$$

or

$$A_W\theta_{i|1}^{\alpha}\theta_{j|1}^{\alpha} + A_\varepsilon(1-\theta_{i|1}^2)^{\alpha/2}\theta_{j|1}^{\alpha} < A_W\theta_{j|1}^{\alpha}\theta_{i|1}^{\alpha} + A_\varepsilon(1-\theta_{j|1}^2)^{\alpha/2}\theta_{i|1}^{\alpha}, \tag{4.43}$$

$$\theta_{i|1} < \theta_{j|1}. \tag{4.44}$$

Combining the results from (4.36) and (4.32), the result follows.

4.10 Proof of Proposition 3.3.3.2

Given the reasoning in Oh and Patton (2015), it follows that as $s \to -\infty$,

$$f_W(s) = \alpha A_W(-s)^{-\alpha-1}, \tag{4.45}$$

where $f_W(s)$ is the probability density of W. Then, using the fact that the tail index α equals the degrees of freedom ν for the Student-t distribution, using Mathematica, it follows that

$$\begin{aligned} A_W &= \lim_{s\to-\infty} \frac{f_W(s)}{\nu(-s)^{-\nu-1}}, \\ &= \frac{(\nu\sigma^2)^{\frac{\nu+1}{2}}}{\nu^{3/2}\sigma B(n/2,1/2)}, \end{aligned}$$

where $B(\cdot,\cdot)$ is the beta function; A_ε directly follows from A_W.

Bibliography

Abramowitz, M. and Stegun, I.A. (1965): Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables. *Dover Publications*.

Alexander, C.O. (2001): Orthogonal GARCH, in "Mastering Risk", Volume II. *FT-Prentice Hall*, 21–38.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999): "Coherent Measures of Risk", *Mathematical Finance*, 9(3), 203–228.

Alexander, G., Baptista, A. (2002): "Economic implications of using a mean-VaR model for portfolio selection: A comparison with mean-variance analysis", *Journal of Economic Dynamics and Control*, **26**, 1159–1193.

Barrodale, I. and Roberts, F. (1974): "Solution of an Overdetermined System of Equations in the L_1 Norm", *Communications of the ACM*, **17**, 319-320.

Barndorff-Nielsen, O. E. (1997): "Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling" *Scandinavian Journal of Statistics*, **24(1)**, 1–13

Barndorff-Nielsen, O. E. and Blaesild, P. (1981): "Hyperbolic Distributions and Ramifications: Contributions to Theory and Applications", in C. Taillie, G. Patil, B. Baldessari (eds.) *Statistical Distributions in Scientific Work*, Volume 4, Reidel, Dordrecht, 19–44.

Bassett, G. W., Koenker, R., Kordas, G. (2004): "Pessimistic Portfolio Allocation and Choquet Expected Utility", *Journal of Financial Econometrics*, **2(4)**, 477-492.

Belloni, A., Chernozhukov, V. (2011): "L_1-penalized Quantile Regression in High-Dimensional Sparse Models", *Annals of Statistics*, **39(1)**, 82–130.

Bera, A. K. and Lee, S. (1993): "Information Matrix Test, Parameter Heterogeneity and ARCH: A Synthesis", *Review of Economic Studies*, **60**, 229–240

Bera, A. K. and Zuo, X.-L. (1996): "Specification test for a linear regression model with ARCH process", *Journal of Statistical Planning and Inference*, **50**, 283–308

Blaesild, P. (1981): "The Two-Dimensional Hyperbolic Distribution and Related Distributions, with an Application to Johannsen's Bean Data", *Biometrika*, **68(1)**, 251–263

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, I. Loris (2009): "Sparse and Stable Markowitz Portfolios", Proceedings of the National Academy of Sciences of the USA, 106, 12267-12272.

Bühlmann, P. and van de Geer, S. (2011): Statistics for High-Dimensional Data. *Springer*.

Chang, C. and Tsay, R. (2010): "Estimation of Covariance Matrix via the Sparse Cholesky Factor with Lasso", *Journal of Statistical Planning and Inference*, **140(12)**, 3858–3873.

Cvitanić, J., Lazrak, A., Martellini, L., and Zapatero, F. (2003): "Optimal Allocation to Hedge Funds: an Empirical Analysis" *Quantitative Finance*, **3(1)**, 28–39

Doornik, J. A. and Hansen, H. (2008): "An omnibus test for univariate and multivariate normality" *Oxford Bulletin of Economics and Statistics*, **70**, 927–939

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation", *Econometrica*, **50(4)**, 987–1007.

Engle, R. and Kroner, K. (1995): "Multivariate Simultaneous GARCH", *Econometric Theory*, **11**, 122–150.

Engle, R. (2002): "Dynamic Conditional Correlation: a Simple Class of Multivariate GARCH Models", *Journal of Business and Economic Statistics*, **20(3)**, 339–350.

Fan, J. and Li, R. (2001): "Variable selection via nonconcave penalized likelihood and its oracle properties", *Journal of American Statistical Association*, **96**, 1348–1360

Fan, J., Y. Zhang and Ke Yu. (2009): "Asset Allocation and Risk Assessment with Gross Exposure Constraints for Vast Portfolios", Working paper, Princeton University.

Fan, J., Han, F. and Liu H. (2013): "Challenges of Big Data analysis", *National Science Review*, **1**, 293–314

Favre, L. and Galeano, J.-A. (2002): "Mean-Modified Value-at-Risk Optimization with Hedge Funds" *The Journal of Alternative Investments*, **5(2)**, 21–25 .

Gaivoronski, A. and Pflug, G. (Winter 2004-2005): "Value-at-risk in Portfolio Optimization: Properties and Computational Approach". *Journal of Risk*, **7(2)**, 1–31.

Ghalanos, A., Rossi, E. and Urga, G. (2015): "Independent Factor Autoregressive Conditional Density Model". *Econometric Reviews*, **34(5)**, 594–616.

Giamouridis, D. and Vrontos, I.D. (2007): "Hedge Fund Portfolio Construction: a Comparison of Static and Dynamic Approaches" *Journal of Banking and Finance*, **31**, 199–217.

Hastie, T., Tibshirani, R., Friedman, J. (2011): "The Elements of Statistical Learning. Data Mining, Inference and Prediction". *Springer*.

Härdle, W. K., Nasekin, S., Lee, D. K. C. and Phoon., K. F. (2015): "TEDAS - Tail Event Driven Asset Allocation". *SFB 649 Discussion Paper 2014-032*, Humboldt-Universität zu Berlin, 2014.

Härdle, W. K., Lee, D. K. C., Nasekin, S., Ni, X. and Petukhina, A. (2015): "Tail Event Driven Asset allocation: evidence from equity and mutual funds' markets". *SFB 649 Discussion Paper 2015-045*, Humboldt-Universität zu Berlin, 2015.

Henze, N. and Zirkler, B. (1990): "A class of invariant consistent tests for multivariate normality", *Communications in Statistics, Theory and Methods*, **19**, 3595–3617.

Hyvärinen, A. and Karhunen, J., Oja, E. (2001): Independent Component Analysis. *John Wiley & Sons*.

Jarque, C. M. and Bera, A. K. (1980): "Efficient tests for normality, homoscedasticity and serial independence of regression residuals", *Economics Letters*, **6**, 255–259.

Jensen, M. B and Lunde, A. (2001): "The NIG-S&ARCH Model: a Fat-Tailed, Stochastic, and Autoregressive Conditional Heteroskedastic Volatility Model". *Econometrics Journal*, **4**, 319–342

Jondeau, E. and Rockinger, M. (2009): "Conditional volatility, skewness and kurtosis: existence, persistence and comovements". *Journal of Economic Dynamics and Control*, **27**, 1699–1737

Jondeau, E. and Rockinger, M. (2009): "The Impact of Shocks on Higher Moments". *Journal of Financial Econometrics*, **7(2)**, 77–105

Jurczenko, E. and Maillet, B. (2006): "Theoretical Foundations of Asset Allocation and Pricing Models with Higher-order Moments", in Jurczenko, E. and Maillet, B. (eds.) *Multi-Moment Asset Allocation and Pricing Models*, Wiley, 1–37

Koenker, R. and Bassett, G. (1978): "Regression Quantiles", *Econometrica*, **46**, 33-50

Lhabitant, F.-S. and Learned, M. (2002): "Hedge Fund Diversification: How Much Is Enough". *The Journal of Alternative Investments*, **5(3)**, 23–49

Li, Y. and Zhu, J. (2008): "L_1-Norm Quantile Regression". *Journal of Computational and Graphical Statistics*, **17(1)**, 1–23

Lintner, J. (1983): The Potential Role of Managed Commodity-Financial Futures Accounts (and/or Funds) in Portfolios of Stocks and Bonds. *Presentation to the Financial Analysts Federation*.

Mardia, K. V. (1970): "Measures of multivariate skewness and kurtosis with applications" *Biometrika*, **57**, 519–530

McFall Lamm, R. (1999): "Portfolios of Alternative Assets: Why Not 100% Hedge Funds"? *Journal of Investing*, **8(4)**, 87–97

McFall Lamm, R. (2003): "Asymmetric Returns and Optimal Hedge Fund Portfolios". *The Journal of Alternative Investments*, **6(2)**, 9–21

Osborne, M.R., Presnell, B. and Turlach, B.A. (2000): "On the LASSO and Its Dual". *Journal of Computational and Graphical Statistics*, **9(2)**, 319–337

Rockafellar, R.T. and Uryasev, S. (2000): "Optimization of Conditional Value-at-Risk" *Journal of Risk*, 2, 21–41

Schmidt, R, Hrycej, T. and Stützle, E. (2006): "Multivariate Distribution Models with Generalized Hyperbolic Margins". *Computational Statistics and Data Analysis*, **50**, 2065–2096

Silverman, B.W. (1998): Density Estimation for Statistics and Data Analysis. *Chapman & Hall/CRC*

Tibshirani, R. (1996): "Regression Shrinkage and Selection via the Lasso". *Journal of Royal Statistical Society*, **58(1)**, 267–288

Wu, L., Yang, Y. and Liu, H. (2014): "Nonnegative-lasso and application in index tracking". *Computational Statistics and Data Analysis*, **70**, 116–126.

Zhang, Y. (1998): "Solving Large-Scale Linear Programs by Interior-Point Methods under the Matlab Environment", *Optimization Methods and Software*, **10**, 1–31

Zheng, Qi, Gallagher, C. and Kulasekera, K.B. (2013): "Adaptive Penalized Quantile Regression for High-Dimensional Data". *Journal of Statistical Planning and Inference*, **143**, 1029–1038

Zou H. (2006): "The Adaptive Lasso and Its Oracle Properties". *Journal of the American Statistical Association*, **101(476)**, 1418–1429

Zou H. and Hastie T. (2005): "Regularization and Variable Selection via the Elastic Net". *Journal of the Royal Statistical Society: Series B*, **67**, 301–320

Aït-Sahalia, Y., Bickel, P. J., Stoker, T. M. (2001): "Goodness-of-fit tests for kernel regression with an application to option implied volatilities", *Journal of Econometrics*, **105(2)**, 363–412

de Boor, C. (2001): "A Practical Guide to Splines. Revised Edition", *Springer*

Berument, H., Kiymaz, H. (2001): "The day of the week effect on stock market volatility", *Journal of Economics and Finance*, **25(2)**, 181–193

Charles, A. (2010): "The day-of-the-week effects on the volatility: The role of the asymmetry", *European Journal of Operational Research*, **202(1)**, 143–152

Cont, R., da Fonseca, J. (2002): "Dynamics of implied volatility surfaces", *Quantitative Finance*, **2(1)**, 45–60

Dragulescu, A. and Yakovenko, V. (2002): "Probability distribution of returns in the Heston model with stochastic volatility", *Quantitative Finance*, **2**, 443–453

Fengler, M. R., Härdle, W. K., Villa, C. (2003): "The dynamics of implied volatilities: a common principal components approach", *Review of Derivatives Research*, **6**, 179–202

Fengler, M. R., Härdle, W. K., & Mammen, E. (2007): "A semiparametric factor model for implied volatility surface dynamics", *Journal of Financial Econometrics*, **5(2)**, 189–218

Franke, J., Härdle, W., & Hafner, C. (2015): "Statistics of Financial Markets: An Introduction", *Springer, 4th edition*

Hall, P., Horowitz, J. (2013): "A simple bootstrap method for constructing nonparametric confidence bands for functions", *The Annals of Statistics*, **41(4)**, 1892–1921

Härdle, W. (1989): "Asymptotic maximal deviation of *M*-Smoothers", *Journal of Multivariate Analysis*, **29**, 163–179

Härdle, W. (1990): "Applied Nonparametric Regression", *Cambridge University Press*

Härdle, W. K., Ritov, Y., Wang, W. (2015): "Tie the straps: uniform bootstrap confidence bands for semiparametric additive models", *Journal of Multivariate Analysis*, **134**, 129–145

Heston, S. L. (1993): "A closed-form solution for options with stochastic volatility with applications to bond and currency options", *The Review of Financial Studies*, **6**, 327–343

Huber, P. J. (1964): "Robust estimation of a location parameter", *Annals of Mathematical Statistics*, **35(1)**, 73–101

Hull, J. C. and White, A. (1987): "The pricing of options on assets with stochastic volatilities", *Journal of Finance*, **42**, 281–300

Lakonishok, J., Smidt, S. (1988): "Are seasonal anomalies real? A ninety-year perspective", *The Review of Financial Studies*, **1(4)**, 403–425

Leung, T., Sircar, R. (2015): "Implied volatility of leveraged ETF options", *Applied Mathematical Finance*, **22(2)**, 162–188

Park, B. U., Mammen, E., Härdle, W., Borak, S. (2009): "Time Series Modelling With Semiparametric Factor Dynamics", *Journal of the American Statistical Association*, **104(485)**, 284–298

Rouah, F. (2013): "The Heston model and its extensions in Matlab and C#", *Wiley*

Ruppert, D., Wand, M. P. (1994): "Multivariate locally weighted least squares regression", *The annals of statistics*, **22(3)**, 1346–1370

Schöbel, R. and Zhu, J. (1999): "Stochastic Volatility With an Ornstein-Uhlenbeck Process: An Extension", *European Finance Review*, **3**, 24–46

van der Stoep, A. W., Grzelak, L. A., Oosterlee, C. W. (2014): "The Heston stochastic-local volatility model: efficient Monte-Carlo simulation", *International Journal of Theoretical and Applied Finance*, **17(7)**, 1–30

Zellner, A., Keuzenkamp, H. A., McAleer, M. (2002): "Simplicity, inference and modelling", *Cambridge University Press*

Basel Committee on Banking Supervision (2013). Global systemically important banks: updated assessment methodology and the higher loss absorbency requirement

Borak, S., Misiorek, A., Weron, R. (2010): "Models for heavy-tailed asset returns", SFB 649 Discussion Papers SFB649DP2010-049, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany

Brechmann, E.C., Hendrich, K., Czado, C. (2013): "Conditional copula simulation for systemic risk stress testing", *Insurance: Mathematics and Economics*, **53**, 722–732

Broda, S. A., Paolella, M. S. (2011): "Expected shortfall for distributions in finance", in: "Statistical Tools for Finance and Insurance", ed. by Cizek, P et al., *Springer*

Chan-Lau, J.A., Chuang, C., Duan, J.-C., Sun, W. (2016): "Banking Network and Systemic Risk via Forward-Looking Partial Default Correlations", research paper of the National University of Singapore Risk Management Institute

Cont, R., Moussa, A. (2010): "Too interconnected to fail: contagion and systemic risk in financial networks", Columbia University - Financial Engineering Report 03

Cont, R., Moussa, A., Santos, E.B. (2013): "Network structure and systemic risk in banking systems", in: Fouque, J.-P., Langsam, J. (Eds.), Handbook on Systemic Risk. *Cambridge University Press*, Cambridge

Feller, W. (1971): "An Introduction to Probability Theory and Its Applications, vol. II", *Wiley*, New York

Fruchterman, T. M. J., Reingold, E. M. (1991): "Graph drawing by force-directed placement", *Journal of Software: Practice and Experience*, **21(11)**, 1129–1164

Girardi, G., Tolga Ergün, A. (2013): "Multivariate GARCH estimation of CoVaR", *Journal of Banking and Finance*, **37**, 3169–3180

Hyung, N., de Vries, C. G. (2007): "Portfolio selection with heavy tails", *Journal of Empirical Finance*, **14**, 383–400

Joe, H. (2015): "Dependence Modeling with Copulas", *CRC Press*

Kalemanova, A., Schmid, B., Werner, R. (2007): "The Normal Inverse Gaussian Distribution for Synthetic CDO Pricing", *The Journal of Derivatives*, **120**, 85–101

McNeil, A. J., Frey, R., Embrechts, P. (2015): "Quantitative Risk Management: Concepts, Techniques and Tools. Revised Edition", *Princeton University Press*

Krupskii, P., Joe, H. (2013): "Factor copula models for multivariate data", *Journal of Multivariate Analysis*, **120**, 85–101

Ng, S. (2006): "Testing cross-section correlation in panel data using spacings", *Journal of Business and Economic Statistics*, **24(1)**, 12–23

Nelsen, R. B. (2006): "An introduction to copulas", *Springer*

Oh, D. H., Patton, A. J. (2015): "Modelling Dependence in High Dimensions with Factor Copulas", *Journal of Business and Economic Statistics*, DOI: 10.1080/07350015.2015.1062384

Oh, D. H., Patton, A. J. (2016): "Time-Varying Systemic Risk: Evidence from a Dynamic Copula Model of CDS Spreads", *Journal of Business and Economic Statistics*, http://dx.doi.org/10.1080/07350015.2016.1177535

Schmidt, R., Stadtmüller, A. (2006): "Non-parametric Estimation of Tail Dependence", *Scandinavian Journal of Statistics*, http://dx.doi.org/10.1111/j.1467-9469.2005.00483.x